



UNIVERSITY OF  
CAMBRIDGE

# Cambridge Working Papers in Economics

*Common Correlated Effects Estimation of  
Heterogeneous Dynamic Panel Data  
Models with Weakly Exogenous Regressors*

*Professor Hashem Pesaran*

*Alexander Chudik*

CWPE 1317

# Common Correlated Effects Estimation of Heterogeneous Dynamic Panel Data Models with Weakly Exogenous Regressors\*

Alexander Chudik<sup>†</sup>

Federal Reserve Bank of Dallas, CAFE and CIMF

M. Hashem Pesaran<sup>‡</sup>

University of Southern California, CAFE, USA, and Trinity College, Cambridge, UK

April 2013

## Abstract

This paper extends the Common Correlated Effects (CCE) approach developed by Pesaran (2006) to heterogeneous panel data models with lagged dependent variable and/or weakly exogenous regressors. We show that the CCE mean group estimator continues to be valid but the following two conditions must be satisfied to deal with the dynamics: a sufficient number of lags of cross section averages must be included in individual equations of the panel, and the number of cross section averages must be at least as large as the number of unobserved common factors. We establish consistency rates, derive the asymptotic distribution, suggest using covariates to deal with the effects of multiple unobserved common factors, and consider jackknife and recursive de-meaning bias correction procedures to mitigate the small sample time series bias. Theoretical findings are accompanied by extensive Monte Carlo experiments, which show that the proposed estimators perform well so long as the time series dimension of the panel is sufficiently large.

**Keywords:** Large panels, lagged dependent variable, cross sectional dependence, coefficient heterogeneity, estimation and inference, common correlated effects, unobserved common factors.

**JEL Classification:** C31, C33.

---

\*We are grateful to Ron Smith, Vanessa Smith, Takashi Yamagata and Qiankun Zhou for helpful comments. In writing of this paper, Chudik benefited from the visit to the Center for Applied Financial Economics (CAFE). Pesaran acknowledges financial support from ESRC grant no. ES/I031626/1.

<sup>†</sup>Federal Reserve Bank of Dallas, 2200 N. Pearl Street, Dallas, Texas. E-mail: alexander.chudik@dal.frb.org. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of Dallas or the Federal Reserve System.

<sup>‡</sup>Department of Economics, University of Southern California, 3620 South Vermont Ave, Los Angeles, California 90089, USA. Email: pesaran@usc.edu; <http://www.econ.cam.ac.uk/faculty/pesaran/>

# 1 Introduction

In a recent paper, Pesaran (2006) proposed the Common Correlated Effects (CCE) approach to estimation of panel data models with multi-factor error structure, which has been further developed by Kapetanios, Pesaran, and Yagamata (2011), Pesaran and Tosetti (2011), and Chudik, Pesaran, and Tosetti (2011). The CCE method is shown to be robust to different types of cross section dependence of errors, possible unit roots in factors, and slope heterogeneity. However, the CCE approach as it was originally proposed does not cover the case where the panel includes a lagged dependent variable and/or weakly exogenous variables as regressors.<sup>1</sup> This paper extends the CCE approach to allow for such regressors. This extension is not straightforward because coefficient heterogeneity in the lags of the dependent variable introduces infinite order lag polynomials in the large  $N$  relationships between cross-sectional averages and the unobserved factors (Chudik and Pesaran, 2013a). Our focus is on stationary heterogeneous panels with weakly exogenous regressors where the cross-sectional dimension ( $N$ ) and the time series dimension ( $T$ ) are sufficiently large. We focus on estimation and inference of the mean coefficients, and consider the application of bias correction techniques to deal with the small  $T$  bias of the estimators.

Recent literature on large dynamic panels focuses mostly on how to deal with cross-sectional (CS) dependence assuming slope homogeneity. Estimation of panel data models with lagged dependent variables and cross-sectionally dependent errors has been considered in Moon and Weidner (2010a and 2010b), who propose a Gaussian quasi maximum likelihood estimator (QMLE).<sup>2</sup> Moon and Weidner's analysis assumes homogeneous coefficients, and therefore is not applicable to dynamic panels with heterogeneous coefficients.<sup>3</sup> Similarly, the interactive-effects estimator (IFE) developed by Bai (2009) also allows for cross-sectionally dependent errors, but assumes homogeneous slopes.<sup>4</sup> Song (2013) extends the analysis of Bai (2009) by allowing for a lagged dependent variable as well as coefficient heterogeneity, but provides results on the estimation of cross-section specific coefficients only. This paper provides an alternative CCE type estimation approach to Song's extension of the

---

<sup>1</sup>See Everaert and Groote (2012) who derive asymptotic bias of CCE pooled estimators in the case of dynamic homogeneous panels.

<sup>2</sup>See also Lee, Moon, and Weidner (2011) for an extension of this framework to panels with measurement errors.

<sup>3</sup>Pesaran and Smith (1995) show that in the presence of coefficient heterogeneity pooled estimators are inconsistent in the case of panel data models with lagged dependent variables.

<sup>4</sup>Earlier literature on large panels typically ignores cross section dependence of errors, including pooled mean group estimation proposed by Pesaran, Shin, and Smith (1999), fully modified OLS estimation by Pedroni (2000) or the panel dynamic OLS estimation by Mark and Sul (2003). These papers can also handle panels with nonstationary data. There is also a large literature on dynamic panels with large  $N$  but finite  $T$ , which assumes homogeneous slopes.

IFE estimator. In addition, we propose a mean group estimator of the mean coefficients, and show that CCE types estimators once augmented with a sufficient number of lags and cross-sectional averages perform well even in the case of dynamic models with weakly exogenous regressors. We also show that the asymptotic distribution of the CCE estimators developed in the literature continue to be applicable to the more general setting considered in this paper. Our method could extend to Song's IFE and we also investigate the performance of the mean group estimator based on Song's unit-specific coefficient estimates.

More specifically, in this paper we considered estimation of autoregressive distributed lagged (ARDL) panel data models where the dependent variable of the  $i^{th}$  cross section unit at time  $t$ ,  $y_{it}$ , is explained by its lagged values, current and lagged values of  $k$  weakly exogenous regressors,  $\mathbf{x}_{it}$ ,  $m$  unobserved (possibly serially correlated) common factors,  $\mathbf{f}_t$ , and a serially uncorrelated idiosyncratic error. In addition to the regressors included in the panel ARDL model, following Pesaran, Smith, and Yamagata (2013) we also assume that there exists a set of additional covariates,  $\mathbf{g}_{it}$ , that are affected by the same set of unobserved common factors,  $\mathbf{f}_t$ . This seems reasonable considering that agents in making their decisions face a common set of factors such as technology, institutional set ups and general economic conditions, which then get manifested in many variables, whether included in the panel data model under consideration or not. Similar arguments also underlie forecasting using a large number of regressors popularized recently in econometrics by Stock and Watson (2002) and Forni et al. (2005).

A necessary condition for the CCE mean group (CCEMG) estimator to be valid in the case of ARDL panel data models is that the number of cross-sectional averages based on  $\mathbf{x}_{it}$  and  $\mathbf{g}_{it}$  must be at least as large as the number of unobserved common factors minus one ( $m - 1$ ). In practice, where the number of unobserved factors is unknown, it is sufficient to assume that the number of available cross-sectional averages is at least  $m_{\max} - 1$ , where  $m_{\max}$  denotes the assumed maximum number of unobserved factors. In most economic applications  $m_{\max}$  is likely to be relatively small.<sup>5</sup>

We also report on the small sample properties of CCEMG estimators for panel ARDL models, using a comprehensive set of Monte Carlo experiments. In particular, we investigate two bias correction methods, namely the half-panel jackknife due to Dhaene and Jochmans, 2012, and the recursive mean adjustment due to So and Shin, 1999. We find that the proposed estimators have

---

<sup>5</sup>Stock and Watson (2002), Giannone, Reichlin, and Sala (2005) conclude that only few, perhaps two, factors explain much of the predictable variations, while Bai and Ng (2007) estimate four factors and Stock and Watson (2005) estimate as many as seven factors.

satisfactory performance under different dynamic parameter configurations, and regardless of the number of unobserved factors, so long as they do not exceed the number of cross-sectional averages, and the time dimension is sufficiently large. We compare the performance of CCEMG with the mean group estimator based on Song’s IFE, and also with Moon and Weidner’s QMLE, Bai’s IFE estimators developed for slope homogeneous ARDL panels. We find that jackknife bias correction is more effective in dealing with the small sample bias than the recursive mean adjustment procedure. Also, the bias correction seems to be helpful only for the coefficients of the lagged dependent variable. The uncorrected CCEMG estimators of the coefficients of the regressors,  $\mathbf{x}_{it}$ , seem to work fine even in the case of panels with a relatively small time dimension.

The remainder of the paper is organized as follows. Section 2 extends the multifactor residual panel data model considered in Pesaran (2006) by introducing lagged dependent variables and allowing the regressors to be weakly exogenous. Section 3 develops a dynamic version of the CCEMG estimator for panel ARDL models. Section 4 discusses the jackknife and recursive demeaning bias correction procedures. Section 5 introduces the mean group estimator based on Song’s individual estimates, describes the Monte Carlo experiments, and reports the small sample results. Mathematical proofs are provided in the Appendix and additional Monte Carlo findings are provided in a Supplement.

## 2 Panel ARDL Model with a Multifactor Error Structure

Suppose that the dependent variable,  $y_{it}$ , the regressors,  $\mathbf{x}_{it}$ , and the covariates,  $\mathbf{g}_{it}$ , are generated according to the following linear covariance stationary dynamic heterogenous panel data model,

$$y_{it} = c_{yi} + \phi_i y_{i,t-1} + \beta'_{0i} \mathbf{x}_{it} + \beta'_{1i} \mathbf{x}_{i,t-1} + u_{it}, \quad (1)$$

$$u_{it} = \gamma'_i \mathbf{f}_t + \varepsilon_{it}, \quad (2)$$

and

$$\boldsymbol{\omega}_{it} = \begin{pmatrix} \mathbf{x}_{it} \\ \mathbf{g}_{it} \end{pmatrix} = \mathbf{c}_{\omega i} + \boldsymbol{\alpha}_i y_{i,t-1} + \boldsymbol{\Gamma}'_i \mathbf{f}_t + \mathbf{v}_{it}, \quad (3)$$

for  $i = 1, 2, \dots, N$  and  $t = 1, 2, \dots, T$ , where  $c_{yi}$  and  $\mathbf{c}_{\omega i}$  are individual fixed effects for unit  $i$ ,  $\mathbf{x}_{it}$  is  $k_x \times 1$  vector of regressors specific to cross-section unit  $i$  at time  $t$ ,  $\mathbf{g}_{it}$  is  $k_g \times 1$  vector of covariates

specific to unit  $i$ ,  $k_x + k_g = k$ ,  $\mathbf{f}_t$  is an  $m \times 1$  vector of unobserved common factors,  $\varepsilon_{it}$  are the idiosyncratic errors,  $\mathbf{\Gamma}_i$  is an  $m \times k$  matrix of factor loadings,  $\boldsymbol{\alpha}_i$  is a  $k \times 1$  vector of unknown coefficients, and  $\mathbf{v}_{it}$  is assumed to follow a general linear covariance stationary process distributed independently of the idiosyncratic errors,  $\varepsilon_{it}$ .

The process for the exogenous variables, (3), can also be written equivalently as a panel ARDL model in  $\boldsymbol{\omega}_{it}$ . But we have chosen to work with this particular specification as it allows us to distinguish between cases of strict and weak exogenous regressors in terms of the feed-back coefficients,  $\boldsymbol{\alpha}_i$ . The case of strictly exogenous regressors, covered in Pesaran (2006), refers to the special case when  $\boldsymbol{\alpha}_i = \mathbf{0}_{k \times 1}$ . As in the earlier literature, the above specification also allows the regressors to be correlated with the unobserved common factors. Lags of  $\mathbf{x}_{it}$  and  $\mathbf{g}_{it}$  are not included in (3), but they could be readily included. In order to keep the notations and exposition simple we also abstract from observed common effects, additional lags of the dependent variable, and other deterministic terms in (1) and (3). Such additional regressors can be readily accommodated at the cost of further notational complexity.

In the above ARDL formulation, we specify the same lag orders for  $y_{it}$  and  $\mathbf{x}_{it}$  because it is desirable in empirical applications to start with a balanced lag order to avoid potential problems connected with persistent regressors. It is also worth noting that a number of panel data models investigated in the literature can be derived as special cases of (1)-(3). The analysis of Moon and Weidner (2010a and 2010b) assumes that  $\beta_{i0} = \beta_0$ ,  $\beta_{i1} = \beta_1$  and  $\phi_i = \phi$ . Bai (2009) assumes  $\beta_{i0} = \beta_0$ ,  $\beta_{i1} = \beta_1$  and  $\phi_i = 0$ . Under the restriction

$$\beta_{1i} = -\phi_i \beta_{0i}, \quad (4)$$

we have

$$y_{it} - \boldsymbol{\theta}'_i \mathbf{x}_{it} = c_{yi} + \phi_i (y_{i,t-1} - \boldsymbol{\theta}'_i \mathbf{x}_{i,t-1}) + u_{it},$$

where  $\boldsymbol{\theta}_i = -\beta_{i1}/\phi_i$ , which in turn can be written as (assuming that  $|\phi_i| < 1$ )

$$y_{it} = c_{yi}^* + \boldsymbol{\theta}'_i \mathbf{x}_{it} + \boldsymbol{\gamma}_i^{*'} \mathbf{f}_t^* + \varepsilon_{it}^*, \quad (5)$$

where  $c_{yi}^* = c_{yi}/(1 - \phi_i)$ ,  $\varepsilon_{it}^* = (1 - \phi_i L)^{-1} \varepsilon_{it}$  is a serially correlated error term, and  $\mathbf{f}_t^*$  is a new set of unobserved common factors. Estimation and inference in panel model (5) have been

studied by Pesaran (2006) who introduced the CCE approach. This approach has been shown to be robust to an unknown number of unobserved common factors (Pesaran, 2006, and Chudik, Pesaran, and Tosetti, 2011), possible unit roots in factors (Kapetanios, Pesaran, and Yagamata, 2011), serial correlation of unknown form in  $\varepsilon_{it}^*$  (Pesaran, 2006), spatial or other forms of weak cross-sectional dependence in  $\varepsilon_{it}^*$  (Pesaran and Tosetti, 2011, and Chudik, Pesaran, and Tosetti, 2011). However, if the restrictions set out in (4) on  $\beta_{0i}$  and  $\beta_{1i}$  do not hold then the CCE approach is no longer applicable and the standard CCE estimators could be seriously biased, even asymptotically.<sup>6</sup> Our objective in this paper is to consider estimation and inference in the panel ARDL model (1)-(3), where the parameter restrictions (4) do not necessarily hold, and the slope coefficients  $\pi_i = (\phi_i, \beta'_{i0}, \beta'_{i1})'$  are allowed to vary across units.

For future reference, partition matrix  $\Gamma_i = (\Gamma_{xi}, \Gamma_{gi})$  into  $m \times k_x$  and  $m \times k_g$  matrices  $\Gamma_{xi}$  and  $\Gamma_{gi}$ , vector  $\alpha_i = (\alpha'_{xi}, \alpha'_{gi})'$  into  $k_x \times 1$  and  $k_g \times 1$  vectors  $\alpha_{xi}$  and  $\alpha_{gi}$ , and similarly  $\mathbf{v}_{it} = (\mathbf{v}'_{xit}, \mathbf{v}'_{git})'$  into  $k_x \times 1$  and  $k_g \times 1$  vectors  $\mathbf{v}_{xit}$  and  $\mathbf{v}_{git}$ .

### 3 Estimation

Let  $\mathbf{z}_{it} = (y_{it}, \mathbf{x}'_{it}, \mathbf{g}'_{it})'$  and write (1)-(3) compactly as

$$\mathbf{A}_{0i}\mathbf{z}_{it} = \mathbf{c}_i + \mathbf{A}_{1i}\mathbf{z}_{i,t-1} + \mathbf{C}_i\mathbf{f}_t + \mathbf{e}_{it}, \quad (6)$$

where  $\mathbf{c}_i = (c_{yi}, \mathbf{c}'_{\omega i})'$ ,  $\mathbf{C}_i = (\gamma_i, \Gamma_i)'$ ,

$$\mathbf{A}_{0i} = \begin{pmatrix} 1 & -\beta'_{0i} & \mathbf{0} \\ \mathbf{0}_{k_x \times 1} & \mathbf{I}_{k_x} & \mathbf{0}_{k_x \times k_g} \\ \mathbf{0}_{k_g \times 1} & \mathbf{0}_{k_g \times k_x} & \mathbf{I}_{k_g} \end{pmatrix}, \quad \mathbf{A}_{1i} = \begin{pmatrix} \phi_i & \beta'_{1i} & \mathbf{0}_{1 \times k_g} \\ \alpha_{xi} & \mathbf{0}_{k_x \times k_x} & \mathbf{0}_{k_x \times k_g} \\ \alpha_{gi} & \mathbf{0}_{k_g \times k_x} & \mathbf{0}_{k_g \times k_g} \end{pmatrix},$$

and  $\mathbf{e}_{it} = (\varepsilon_{it}, \mathbf{v}'_{it})'$  is a serially correlated error process.  $\mathbf{A}_{0i}$  is invertible (for any  $i$ ) and multiplying (6) by  $\mathbf{A}_{0i}^{-1}$ , we obtain the following reduced form VAR(1) representation of  $\mathbf{z}_{it}$  with serially correlated errors,

$$\mathbf{z}_{it} = \mathbf{c}_{zi} + \mathbf{A}_i\mathbf{z}_{i,t-1} + \mathbf{A}_{0i}^{-1}\mathbf{C}_i\mathbf{f}_t + \mathbf{e}_{zit},$$

---

<sup>6</sup>See Everaert and Groote (2012) for derivation of asymptotic bias of CCE pooled estimators in the case of dynamic homogeneous panels.

where  $\mathbf{c}_{zi} = \mathbf{A}_{0i}^{-1} \mathbf{c}_i$ ,  $\mathbf{e}_{zit} = \mathbf{A}_{0i}^{-1} \mathbf{e}_{it}$ , and  $\mathbf{A}_i = \mathbf{A}_{0i}^{-1} \mathbf{A}_{1i}$ .

We postulate the following assumptions for the estimation of the short-run coefficients.

**ASSUMPTION 1** (*Individual Specific Errors*) The individual specific errors  $\varepsilon_{it}$  and  $\mathbf{v}_{jt'}$  are independently distributed for all  $i, j, t$  and  $t'$ . The vector of errors  $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, \dots, \varepsilon_{Nt})'$  is spatially correlated according to

$$\boldsymbol{\varepsilon}_t = \mathbf{R} \boldsymbol{\varsigma}_{\varepsilon t}, \quad (7)$$

where the  $N \times N$  matrix  $\mathbf{R}$  has bounded row and column matrix norms, namely  $\|\mathbf{R}\|_{\infty} < K$  and  $\|\mathbf{R}\|_1 < K$ , respectively, for some constant  $K < \infty$ , which does not depend on  $N$ , diagonal elements of  $\mathbf{R}\mathbf{R}'$  are bounded away from zero,  $\boldsymbol{\varsigma}_{\varepsilon t} = (\varsigma_{\varepsilon 1t}, \varsigma_{\varepsilon 2t}, \dots, \varsigma_{\varepsilon Nt})'$ , and  $\varsigma_{\varepsilon it}$ , for  $i = 1, 2, \dots, N$  and  $t = 1, 2, \dots, T$ , are independently and identically distributed (IID) with mean 0, unit variances, and finite fourth-order moments. For each  $i = 1, 2, \dots, N$ ,  $\mathbf{v}_{it}$  follows a linear stationary process with absolute summable autocovariances (uniformly in  $i$ ),

$$\mathbf{v}_{it} = \sum_{\ell=0}^{\infty} \mathbf{S}_{i\ell} \boldsymbol{\varsigma}_{v,i,t-\ell}, \quad (8)$$

where  $\boldsymbol{\varsigma}_{vit}$  is a  $k \times 1$  vector of IID random variables, with mean zero, variance matrix  $\mathbf{I}_k$  and finite fourth-order moments. In particular,

$$\|Var(\mathbf{v}_{it})\| = \left\| \sum_{\ell=0}^{\infty} \mathbf{S}_{i\ell} \mathbf{S}_{i\ell}' \right\| \leq K < \infty, \quad (9)$$

for  $i = 1, 2, \dots, N$ , where  $\|\mathbf{A}\|$  is the spectral norm of the matrix  $\mathbf{A}$ .

**ASSUMPTION 2** (*Common Effects*) The  $m \times 1$  vector of unobserved common factors,  $\mathbf{f}_t = (f_{1t}, f_{2t}, \dots, f_{mt})'$ , is covariance stationary with absolute summable autocovariances, distributed independently of the individual specific errors  $\varepsilon_{it'}$  and  $\mathbf{v}_{it'}$  for all  $i, t$  and  $t'$ . Fourth moments of  $f_{\ell t}$ , for  $\ell = 1, 2, \dots, m$ , are bounded.

**ASSUMPTION 3** (*Factor Loadings*) The factor loadings  $\boldsymbol{\gamma}_i$ , and  $\boldsymbol{\Gamma}_i$  are independently and identically distributed across  $i$ , and of the common factors  $\mathbf{f}_t$ , for all  $i$  and  $t$ , with mean  $\boldsymbol{\gamma}$  and  $\boldsymbol{\Gamma}$ , respectively, and bounded second moments. In particular,

$$\boldsymbol{\gamma}_i = \boldsymbol{\gamma} + \boldsymbol{\eta}_{\boldsymbol{\gamma}i}, \quad \boldsymbol{\eta}_{\boldsymbol{\gamma}i} \sim IID \left( \mathbf{0}_{m \times 1}, \boldsymbol{\Omega}_{\boldsymbol{\gamma}} \right), \quad \text{for } i = 1, 2, \dots, N,$$



and

$$\text{vec}(\Gamma_i) = \text{vec}(\Gamma) + \boldsymbol{\eta}_{\Gamma i}, \quad \boldsymbol{\eta}_{\Gamma i} \sim IID \left( \mathbf{0}_{km \times 1}, \boldsymbol{\Omega}_{\Gamma} \right), \text{ for } i = 1, 2, \dots, N,$$

where  $\boldsymbol{\Omega}_{\gamma}$  and  $\boldsymbol{\Omega}_{\Gamma}$  are  $m \times m$  and  $km \times km$  symmetric nonnegative definite matrices,  $\|\boldsymbol{\gamma}\| < K$ ,  $\|\boldsymbol{\Omega}_{\gamma}\| < K$ ,  $\|\Gamma\| < K$ , and  $\|\boldsymbol{\Omega}_{\Gamma}\| < K$ .

**ASSUMPTION 4** (*Heterogenous Coefficients*)  $(2k_x + 1) \times 1$  dimensional vector of coefficients  $\boldsymbol{\pi}_i = (\phi_i, \boldsymbol{\beta}'_{0i}, \boldsymbol{\beta}'_{1i})'$  follows the random coefficient model

$$\boldsymbol{\pi}_i = \boldsymbol{\pi} + \mathbf{v}_{\pi i}, \quad \mathbf{v}_{\pi i} \sim IID \left( \mathbf{0}_{2k_x+1 \times 1}, \boldsymbol{\Omega}_{\pi} \right), \text{ for } i = 1, 2, \dots, N, \quad (10)$$

where  $\boldsymbol{\pi} = (\phi, \boldsymbol{\beta}'_0, \boldsymbol{\beta}'_1)'$ ,  $\|\boldsymbol{\pi}\| < K$ ,  $\|\boldsymbol{\Omega}_{\pi}\| < K$ ,  $\boldsymbol{\Omega}_{\pi}$  is  $(2k_x + 1) \times (2k_x + 1)$  symmetric nonnegative definite matrix, and the random deviations  $\mathbf{v}_{\pi i}$  are independently distributed of  $\boldsymbol{\gamma}_j$ ,  $\Gamma_j$ ,  $\varepsilon_{jt}$ ,  $\mathbf{v}_{jt}$ , and  $\mathbf{f}_t$  for all  $i, j$ , and  $t$ . Furthermore, the support of  $\phi_i$  lies strictly inside the unit circle, and  $E \|\mathbf{c}_i\| < K$  for all  $i$ .

**ASSUMPTION 5** (*Regressors and Covariates*) Regressors and covariates in  $\boldsymbol{\omega}_{it} = (\mathbf{x}'_{it}, \mathbf{g}'_{it})'$  are either strictly exogenous and generated according to the canonical factor model (3) with  $\boldsymbol{\alpha}_i = \mathbf{0}_{k \times 1}$ , or weakly exogenous and generated according to (3) with  $\boldsymbol{\alpha}_i$ , for  $i = 1, 2, \dots, N$ , IID across  $i$  and independently distributed of  $\mathbf{v}_{\pi j}$ ,  $\boldsymbol{\gamma}_j$ ,  $\Gamma_j$ ,  $\varepsilon_{jt}$ ,  $\mathbf{v}_{jt}$ , and  $\mathbf{f}_t$  for all  $i, j$  and  $t$ . In the case where the regressors are weakly exogenous we also assume:

- (i) the support of  $\lambda_1(\mathbf{A}_i)$  lies strictly inside the unit circle, for  $i = 1, 2, \dots, N$ , where  $\mathbf{A}_i = \mathbf{A}_{0i}^{-1} \mathbf{A}_{1i}$ , and  $\lambda_1(\mathbf{A}_i)$  denotes the largest eigenvalue (in absolute value) of  $\mathbf{A}_i$ ; and
- (ii) the inverse of polynomial  $\boldsymbol{\Lambda}(L) = \sum_{\ell=0}^{\infty} \boldsymbol{\Lambda}_{\ell} L^{\ell}$ , where  $\boldsymbol{\Lambda}_{\ell} = E(\mathbf{A}_i^{\ell} \mathbf{A}_{0i}^{-1})$ , exists and has exponentially decaying coefficients.

Let  $\mathbf{w} = (w_1, w_2, \dots, w_N)'$  be an  $N \times 1$  vector of non-stochastic (or pre-determined) weights that satisfies the following ‘granularity’ conditions

$$\|\mathbf{w}\| = O\left(N^{-\frac{1}{2}}\right), \quad (11)$$

$$\frac{w_i}{\|\mathbf{w}\|} = O\left(N^{-\frac{1}{2}}\right) \text{ uniformly in } i, \quad (12)$$

and the normalization condition

$$\sum_{i=1}^N w_i = 1. \quad (13)$$

The weights vector  $\mathbf{w}$  depends on  $N$ , but we suppress the subscript  $N$  to simplify notations.

Next, we derive a large  $N$  representation for cross-sectional averages of  $\mathbf{z}_{it}$  following Chudik and Pesaran (2013a). Since the support of the eigenvalues of  $\mathbf{A}_i$  is assumed to lie strictly inside the unit circle,  $\mathbf{z}_{it}$  is an invertible covariance stationary process and can be written as

$$\mathbf{z}_{it} = \sum_{\ell=0}^{\infty} \mathbf{A}_i^{\ell} (\mathbf{c}_{zi} + \mathbf{A}_{0i}^{-1} \mathbf{C}_i \mathbf{f}_{t-\ell} + \mathbf{e}_{z,i,t-\ell}),$$

for  $i = 1, 2, \dots, N$ . Taking weighted cross-sectional averages of the above and making use of the fact that under our assumptions the elements of  $\mathbf{e}_{zit}$  are weakly cross-sectionally dependent, together with the random coefficients Assumptions 3-5, we have

$$\sum_{i=1}^N \sum_{\ell=0}^{\infty} w_i \mathbf{A}_i^{\ell} \mathbf{e}_{z,i,t-\ell} = O_p(N^{-1/2}).$$

Since (under Assumptions 3-5)  $\mathbf{A}_i$  and  $\mathbf{A}_{0,i}$  are independently distributed of  $\mathbf{C}_i$ , and  $\mathbf{A}_i$ ,  $\mathbf{A}_{0,i}$  and  $\mathbf{C}_i$  are independently distributed across  $i$ , we have

$$\begin{aligned} \sum_{i=1}^N \sum_{\ell=0}^{\infty} w_i \mathbf{A}_i^{\ell} \mathbf{A}_{0,i}^{-1} \mathbf{C}_i \mathbf{f}_{t-\ell} &= \sum_{\ell=0}^{\infty} E(\mathbf{A}_i^{\ell} \mathbf{A}_{0,i}^{-1} \mathbf{C}_i) \mathbf{f}_{t-\ell} + O_p(N^{-1/2}), \\ &= \mathbf{\Lambda}(L) \mathbf{C} \mathbf{f}_t + O_p(N^{-1/2}), \end{aligned}$$

where  $\mathbf{C} = E(\mathbf{C}_i) = (\boldsymbol{\gamma}, \boldsymbol{\Gamma})'$ . Thus, yielding the following large  $N$  representation

$$\tilde{\mathbf{z}}_{wt} = \mathbf{\Lambda}(L) \mathbf{C} \mathbf{f}_t + O_p(N^{-1/2}), \quad (14)$$

where  $\tilde{\mathbf{z}}_{wt} = \bar{\mathbf{z}}_{wt} - \bar{\mathbf{c}}_{zw}$  is  $k+1$  dimensional vector of de-trended cross section averages,  $\bar{\mathbf{z}}_{wt} = (\bar{y}_{wt}, \bar{\mathbf{x}}'_{wt}, \bar{\mathbf{g}}'_{wt})' = \sum_{i=1}^N w_i \mathbf{z}_{it}$  is  $k+1$  dimensional vector of cross section averages, and  $\bar{\mathbf{c}}_{zw} = \sum_{i=1}^N w_i (\mathbf{I}_{k+1} - \mathbf{A}_i)^{-1} \mathbf{c}_{zi}$ .

Multiplying (14) by the inverse of  $\mathbf{\Lambda}(L)$  now yields the following large  $N$  expression for a linear

combination of the unobserved common factors,

$$\mathbf{C}\mathbf{f}_t = \mathbf{\Lambda}^{-1}(L) \tilde{\mathbf{z}}_{wt} + O_p(N^{-1/2}). \quad (15)$$

Consider now the special case where  $\boldsymbol{\alpha}_i = \mathbf{0}_{k \times 1}$ , and the regressors are strictly exogenous. In this case the regressors are independently distributed of the coefficients in  $\boldsymbol{\pi}_i = (\phi_i, \boldsymbol{\beta}'_{0,i}, \boldsymbol{\beta}'_{1,i})'$ , which simplifies the derivation of the large  $N$  representation for  $\tilde{\mathbf{z}}_{wt}$ . In particular,  $(1 - \phi_i L)$  is invertible for any  $i = 1, 2, \dots, N$  under Assumption 4, and multiplying (1) by  $(1 - \phi_i L)^{-1}$  we have

$$y_{it} = \sum_{\ell=0}^{\infty} \phi_i^\ell c_{yi} + \sum_{\ell=0}^{\infty} \phi_i^\ell \boldsymbol{\beta}'_{0,i} \mathbf{x}_{i,t-\ell} + \sum_{\ell=0}^{\infty} \phi_i^\ell \boldsymbol{\beta}'_{1,i} \mathbf{x}_{i,t-\ell-1} + \sum_{\ell=0}^{\infty} \phi_i^\ell \boldsymbol{\gamma}'_i \mathbf{f}_{t-\ell} + \sum_{\ell=0}^{\infty} \phi_i^\ell \varepsilon_{i,t-\ell}. \quad (16)$$

Taking weighted cross-sectional averages, under Assumptions 1-5, and assuming  $\boldsymbol{\alpha}_i = \mathbf{0}_{k \times 1}$ , we obtain

$$\bar{y}_{wt} = \bar{c}_{yw} + a(L) \boldsymbol{\gamma}' \mathbf{f}_t + a(L) (\boldsymbol{\beta}'_0 + \boldsymbol{\beta}'_1 L) \bar{\mathbf{x}}_{wt} + O_p(N^{-1/2}), \quad (17)$$

and

$$\bar{\boldsymbol{\omega}}_{wt} = \bar{\mathbf{c}}_{\omega w} + \mathbf{\Gamma}' \mathbf{f}_t + O_p(N^{-1/2}), \quad (18)$$

where  $\bar{c}_{yw} = \sum_{i=1}^N w_i c_{y,i} (1 - \phi_i)^{-1}$ ,  $\bar{\mathbf{c}}_{\omega w} = \sum_{i=1}^N w_i \mathbf{c}_{\omega i}$ , and  $a(L) = \sum_{\ell=0}^{\infty} a_\ell L^\ell$  with its elements given by the moments of  $\phi_i$ , namely  $a_\ell = E(\phi_i^\ell)$ , for  $\ell = 0, 1, 2, \dots$ . Note that under Assumption 4, which constraints the support of  $\phi_i$  to lie strictly inside the unit circle, the rate of decay of the coefficients in  $a(L)$  is exponential. This restriction on the support of  $\phi_i$  also ensures the existence of all moments of  $\phi_i$ . The rate of decay of the coefficients of  $a(L)$  will not necessarily be exponential if the support of  $\phi_i$  covered 1, and depending on the properties of the distribution of  $\phi_i$  in the neighborhood of 1,  $a(L)$  need not be absolute summable, in which case  $\bar{y}_{wt}$  could converge (in a quadratic mean) to a long memory process as  $N \rightarrow \infty$ . Such possibilities are ruled out by Assumption 4.

However, under Assumption 4 and By Lemma A.1 of Chudik and Pesaran (2013b), the inverse of  $a(L)$  exists and has exponentially decaying coefficients. Pre-multiplying both sides of (17) by  $b(L) = a^{-1}(L)$ , we obtain

$$\boldsymbol{\gamma}' \mathbf{f}_t = b(L) \bar{y}_{wt} - b(1) \bar{c}_{yw} - \boldsymbol{\beta}'_0 \bar{\mathbf{x}}_{wt} - \boldsymbol{\beta}'_1 \bar{\mathbf{x}}_{w,t-1} + O_p(N^{-1/2}). \quad (19)$$

Stacking equations (18) and (19), we obtain (15) with  $\mathbf{\Lambda}^{-1}(L)$  reduced (in the strictly exogenous case) to

$$\mathbf{\Lambda}^{-1}(L) = \begin{pmatrix} b(L) & -\beta'_0 - \beta'_1 L & \mathbf{0}_{1 \times k_g} \\ \mathbf{0}_{k_x \times 1} & \mathbf{I}_{k_x} & \mathbf{0}_{k_x \times k_g} \\ \mathbf{0}_{k_g \times 1} & \mathbf{0}_{k_g \times k_x} & \mathbf{I}_{k_g} \end{pmatrix}. \quad (20)$$

It follows from (15) that when  $\text{rank}(\mathbf{C}) = m$  and regardless of whether the regressors are weakly or strictly exogenous, de-trended cross section averages  $\tilde{\mathbf{z}}_{wt}$  and their lags can be used as proxies for the unobserved common factors, assuming that  $N$  is sufficiently large, namely we have

$$\mathbf{f}_t = \mathbf{G}(L) \tilde{\mathbf{z}}_{wt} + O_p\left(N^{-1/2}\right), \quad (21)$$

where

$$\mathbf{G}(L) = (\mathbf{C}'\mathbf{C})^{-1} \mathbf{C}'\mathbf{\Lambda}^{-1}(L).$$

Note that the coefficients of the distributed lag function,  $\mathbf{G}(L)$ , decay at an exponential rate. In particular, in the case of strictly exogenous regressors (where  $\alpha_i = \mathbf{0}_{k \times 1}$ ), the decay rate of the coefficients in  $\mathbf{G}(L)$  is given by the decay rate of the coefficients in  $b(L)$ , see (20) and (23). As established by Lemma A.1 of Chudik and Pesaran (2013b), the decay rate of the coefficients in  $b(L)$  is exponential under Assumption 4, which confines the support of  $\phi_i$  to lie strictly within the unit circle. In the case of weakly exogenous regressors, an exponential rate of decay of the coefficients in  $\mathbf{\Lambda}^{-1}(L)$  is ensured by Assumption 5-ii.

The full column rank of  $\mathbf{C}$  ensures that  $\mathbf{C}'\mathbf{C}$  is invertible and this rank condition is required for the estimation of unit-specific coefficients. In contrast, the rank condition is not always necessary for estimation of the cross-sectional mean of the coefficients, as we shall see below.

**ASSUMPTION 6**  $(k+1) \times m$  dimensional matrix  $\mathbf{C} = (\boldsymbol{\gamma}, \boldsymbol{\Gamma})'$  has full column rank.

Substituting the large  $N$  representation for the unobserved common factors (21) into (1), we obtain

$$y_{it} = c_{yi}^* + \phi_i y_{i,t-1} + \beta'_{0i} \mathbf{x}_{it} + \beta'_{1i} \mathbf{x}_{i,t-1} + \boldsymbol{\delta}'_i(L) \bar{\mathbf{z}}_{wt} + \varepsilon_{it} + O_p\left(N^{-1/2}\right), \quad (22)$$

where

$$\boldsymbol{\delta}_i(L) = \sum_{\ell=0}^{\infty} \boldsymbol{\delta}_{i\ell} L^\ell = \mathbf{G}'(L) \boldsymbol{\gamma}_i, \quad (23)$$

and  $c_{yi}^* = c_{yi} - \boldsymbol{\delta}_i'(1) \bar{\mathbf{c}}_{zw}$ .

Consider now the following cross-sectionally augmented regressions, based on (22),

$$y_{it} = c_{yi}^* + \phi_i y_{i,t-1} + \boldsymbol{\beta}_{0i}' \mathbf{x}_{it} + \boldsymbol{\beta}_{1i}' \mathbf{x}_{i,t-1} + \sum_{\ell=0}^{p_T} \boldsymbol{\delta}_{i\ell}' \bar{\mathbf{z}}_{w,t-\ell} + e_{yit}, \quad (24)$$

where  $p_T$  is the number of lags (assumed to be the same across units, for the simplicity of exposition).

The error term,  $e_{yit}$ , can be decomposed into three parts: an idiosyncratic term,  $\varepsilon_{it}$ , an error component due to the truncation of possibly infinite polynomial distributed lag function,  $\boldsymbol{\delta}_i(L)$ , and an error component due to the approximation of unobserved common factors, namely

$$e_{yit} = \varepsilon_{it} + \sum_{\ell=p_T+1}^{\infty} \boldsymbol{\delta}_{i\ell}' \bar{\mathbf{z}}_{w,t-\ell} + O_p(N^{-1/2}).$$

Note that the coefficients of the distributed lag function,  $\boldsymbol{\delta}_i(L) = \boldsymbol{\gamma}_i' \mathbf{G}(L)$ , decay at an exponential rate.

Let  $\hat{\boldsymbol{\pi}}_i = (\hat{\phi}_i, \hat{\boldsymbol{\beta}}_{0i}', \hat{\boldsymbol{\beta}}_{1i}')'$  be the least squares estimates of  $\boldsymbol{\pi}_i$  based on the cross-sectionally augmented regression (24). Also consider the following data matrices

$$\boldsymbol{\Xi}_i = \begin{pmatrix} y_{ip_T} & \mathbf{x}_{i,p_T+1}' & \mathbf{x}_{ip_T}' \\ y_{i,p_T+1} & \mathbf{x}_{i,p_T+2}' & \mathbf{x}_{i,p_T+1}' \\ \vdots & \vdots & \vdots \\ y_{i,T-1} & \mathbf{x}_{iT}' & \mathbf{x}_{i,T-1}' \end{pmatrix}, \quad \bar{\mathbf{Q}}_w = \begin{pmatrix} 1 & \bar{\mathbf{z}}_{w,p_T+1}' & \bar{\mathbf{z}}_{w,p_T}' & \cdots & \bar{\mathbf{z}}_{w,1}' \\ 1 & \bar{\mathbf{z}}_{w,p_T+2}' & \bar{\mathbf{z}}_{w,p_T+1}' & \cdots & \bar{\mathbf{z}}_{w,2}' \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \bar{\mathbf{z}}_{w,T}' & \bar{\mathbf{z}}_{w,T-1}' & \cdots & \bar{\mathbf{z}}_{w,T-p_T}' \end{pmatrix}, \quad (25)$$

and the projection matrix

$$\bar{\mathbf{M}}_q = \mathbf{I}_{T-p_T} - \bar{\mathbf{Q}}_w (\bar{\mathbf{Q}}_w' \bar{\mathbf{Q}}_w)^+ \bar{\mathbf{Q}}_w',$$

where  $\mathbf{I}_{T-p_T}$  is a  $(T-p_T) \times (T-p_T)$  dimensional identity matrix, and  $\mathbf{A}^+$  denotes the Moore-Penrose generalized inverse of  $\mathbf{A}$ . Matrices  $\boldsymbol{\Xi}_i$ ,  $\bar{\mathbf{Q}}_w$ , and  $\bar{\mathbf{M}}_q$  depend also on  $p_T$ ,  $N$  and  $T$ , but we omit these subscripts to simplify notations. We summarize and introduce additional notations that will be useful (for proofs) in Appendix A.1.

$\hat{\pi}_i$  can now be written as

$$\hat{\pi}_i = (\Xi_i' \bar{\mathbf{M}}_q \Xi_i)^{-1} \Xi_i' \bar{\mathbf{M}}_q \mathbf{y}_i, \quad (26)$$

where  $\mathbf{y}_i = (y_{i,p_T+1}, y_{i,p_T+2}, \dots, y_{i,T})'$ . The mean group estimator of  $\pi = E(\pi_i) = (\phi, \beta'_0, \beta'_1)'$  is given by

$$\hat{\pi}_{MG} = \frac{1}{N} \sum_{i=1}^N \hat{\pi}_i. \quad (27)$$

In addition to Assumptions 1-6 above, we shall also require the following further assumption.

**ASSUMPTION 7** (a) Denote the  $(t - p_T)$ -th row of matrix  $\tilde{\Xi}_i = \mathbf{M}_h \Xi_i$  by  $\tilde{\xi}_{it}' = (\tilde{\xi}_{i1t}, \tilde{\xi}_{i2t}, \dots, \tilde{\xi}_{i,2k_x+1,t})$ , where  $\mathbf{M}_h$  is defined in the Appendix by (A.4). Individual elements of  $\tilde{\xi}_{it}$  have uniformly bounded fourth moments, namely there exists a positive constant  $K$  such that  $E(\tilde{\xi}_{ist}^4) < K$  for any  $t = p_T + 1, p_T + 2, \dots, T$ ,  $i = 1, 2, \dots, N$  and  $s = 1, 2, \dots, 2k_x + 1$ .

(b) There exists  $N_0$  and  $T_0$  such that for all  $N \geq N_0$ ,  $T \geq T_0$ ,  $(2k_x + 1) \times (2k_x + 1)$  matrices  $\hat{\Psi}_{\Xi,iT}^{-1} = (\Xi_i' \bar{\mathbf{M}}_q \Xi_i / T)^{-1}$  exist for all  $i$ .

(c)  $(2k_x + 1) \times (2k_x + 1)$  dimensional matrix  $\Sigma_{i\xi}$  defined in (A.14) in the Appendix is invertible for all  $i$  and  $\|\Sigma_{i\xi}^{-1}\| < K < \infty$  for all  $i$ .

This assumption plays a similar role as Assumption 4.6 in Chudik, Pesaran, and Tosetti (2011) and ensures that  $\hat{\pi}_i$ ,  $\hat{\pi}_{MG}$  and their asymptotic distributions are well defined.

First, we establish sufficient conditions for the consistency of unit-specific estimates.

**Theorem 1 (Consistency of  $\hat{\pi}_i$ )** Suppose  $y_{it}$ , for  $i = 1, 2, \dots, N$  and  $t = 1, 2, \dots, T$  is given by the panel ARDL model (1)-(3), and Assumptions 1-7 hold. Then, as  $(N, T, p_T) \xrightarrow{j} \infty$ , such that  $p_T^3/T \rightarrow \varkappa$ ,  $0 < \varkappa < \infty$ , we have

$$\hat{\pi}_i - \pi_i \xrightarrow{P} \mathbf{0}_{2k_x+1 \times 1}, \quad (28)$$

where  $\hat{\pi}_i = (\hat{\phi}_i, \hat{\beta}'_{0i}, \hat{\beta}'_{1i})'$  is given by (26).

No restrictions on the relative expansion rates of  $N$  and  $T$  to infinity are required for the consistency of  $\hat{\pi}_i$  in the theorem above, but the number of lags needs to be restricted so that there are sufficient degrees of freedom for consistent estimation (i.e. the number of lags is not too large, in particular it is required that  $p_T^2/T \rightarrow 0$ ) and the bias due to the truncation of (possibly) infinite

lag polynomials is sufficiently small (i.e. the number of lags is not too small, in our case  $\sqrt{T}\rho^{p_T} \rightarrow 0$  for some positive constant  $\rho < 1$ ). Letting  $p_T^3/T \rightarrow \varkappa$ ,  $0 < \varkappa < \infty$ , as  $T \rightarrow \infty$ , ensures that these conditions are met.<sup>7</sup> The rank condition in Assumption 6 is also necessary for the consistency of  $\hat{\pi}_i$ . This is because the unobserved factors are allowed to be serially correlated as well as being correlated with the regressors.

### 3.1 Consistency and asymptotic distribution of $\hat{\pi}_{MG}$

Consistency of the unit-specific estimates  $\hat{\pi}_i$  is not always necessary for the consistency of the mean group estimator of  $\pi = E(\pi_i)$ , which is established next.

**Theorem 2 (Consistency of  $\hat{\pi}_{MG}$ )** *Suppose  $y_{it}$ , for  $i = 1, 2, \dots, N$  and  $t = 1, 2, \dots, T$  is given by the panel data model (1)-(3), and Assumptions 1-5 and 7 hold, and  $(N, T, p_T) \xrightarrow{j} \infty$ , such that  $p_T^3/T \rightarrow \varkappa$ ,  $0 < \varkappa < \infty$ . Then,*

(i) *if Assumption 6 also holds,*

$$\hat{\pi}_{MG} - \pi \xrightarrow{p} \mathbf{0}_{2k_x+1 \times 1}, \quad (29)$$

where  $\hat{\pi}_{MG} = (\hat{\phi}_{MG}, \hat{\beta}'_{0MG}, \hat{\beta}'_{1MG})'$  is given by (27);

(ii) *if Assumption 6 does not hold but  $\mathbf{f}_t$  is serially uncorrelated,  $\hat{\pi}_{MG} - \pi \xrightarrow{p} \mathbf{0}_{2k_x+1 \times 1}$ .*

Theorem 2 establishes that  $\hat{\pi}_{MG}$  is consistent (as  $N$  and  $T$  tend jointly to infinity at any rate), regardless of the rank condition when factors are serially uncorrelated, although they can still be correlated with the regressors. When the factors are serially correlated, then the rank condition is required for the consistency of  $\hat{\pi}_{MG}$ . As we have seen, full column rank of  $\mathbf{C}$  is sufficient for approximating the unobserved common factors arbitrarily well by cross section averages and their lags. In this case, the serial correlation of factors and correlation of factors and regressors do not pose any problems. When the rank condition does not hold, but factors are serially uncorrelated, then  $\hat{\pi}_i$  could be inconsistent due to the correlation of  $\mathbf{x}_{it}$  and  $\mathbf{f}_t$ , but the asymptotic bias of  $\hat{\pi}_i - \pi_i$  is cross-sectionally weakly dependent with zero mean and consequently the mean group estimator is consistent.

The following theorem establishes the asymptotic distribution of  $\hat{\pi}_{MG}$ .

---

<sup>7</sup>See also a related discussion in Berk (1974), Chudik and Pesaran (2013b) and Said and Dickey (1984) on the truncation of infinite polynomials in least squares regressions.

**Theorem 3 (Asymptotic distribution of  $\hat{\pi}_{MG}$ )** Suppose  $y_{it}$ , for  $i = 1, 2, \dots, N$  and  $t = 1, 2, \dots, T$  are generated by the panel ARDL model (1)-(3), Assumptions 1-5 and 7 hold, and  $(N, T, p_T) \xrightarrow{j} \infty$  such that  $N/T \rightarrow \kappa_1$  and  $p_T^3/T \rightarrow \kappa_2$ ,  $0 < \kappa_1, \kappa_2 < \infty$ . Then,

(i) if Assumption 6 also holds, we have

$$\sqrt{N}(\hat{\pi}_{MG} - \pi) \xrightarrow{d} N\left(\mathbf{0}_{2k_x+1 \times 1}, \mathbf{\Omega}_\pi\right), \quad (30)$$

(ii) if Assumption 6 does not hold, but  $\mathbf{f}_t$  is serially uncorrelated, we have

$$\sqrt{N}(\hat{\pi}_{MG} - \pi) \xrightarrow{d} N\left(\mathbf{0}_{2k_x+1 \times 1}, \mathbf{\Sigma}_{MG}\right), \quad (31)$$

where  $\hat{\pi}_{MG} = (\hat{\phi}'_{MG}, \hat{\beta}'_{0MG}, \hat{\beta}'_{1MG})'$  is given by (27) and  $\mathbf{\Sigma}_{MG}$  is given by equation (A.84) in the Appendix.

In both cases, the asymptotic variance of  $\hat{\pi}_{MG}$  can be consistently estimated nonparametrically by

$$\hat{\Sigma}_{MG} = \frac{1}{N-1} \sum_{i=1}^N (\hat{\pi}_i - \hat{\pi}_{MG})(\hat{\pi}_i - \hat{\pi}_{MG})'. \quad (32)$$

The convergence rate of  $\hat{\pi}_{MG}$  is  $\sqrt{N}$  due to the heterogeneity of the coefficients. Theorem 3 shows that the asymptotic distribution of  $\hat{\pi}_{MG}$  differs depending on the rank of the matrix  $\mathbf{C}$  in Assumption 6. If  $\mathbf{C}$  has full column rank, then the unit specific estimates  $\hat{\pi}_i$  are consistent,  $\mathbf{\Sigma}_{MG}$  reduces to  $\mathbf{\Omega}_\pi$ , and the asymptotic variance of the mean group estimator is given by the variance of  $\pi_i$  alone. If, on the other hand,  $\mathbf{C}$  does not have the full column rank and factors are serially uncorrelated then the unit-specific estimates are inconsistent (since  $\mathbf{f}_t$  is correlated with  $\mathbf{x}_{it}$ ), but  $\hat{\pi}_{MG}$  is consistent and asymptotically normal with variance that depends not only on  $\mathbf{\Omega}_\pi$  but also on other parameters including the variance of factor loadings. Pesaran (2006) did not require any restrictions on the relative rate of convergence of  $N$  and  $T$  for the asymptotic distribution of the common correlated mean group estimator. This is no longer the case in our model due to  $O(T^{-1})$  time series bias of  $\hat{\pi}_i$  and  $\hat{\pi}_{MG}$  that arises from the presence of lagged values of the dependent variable. This bias dates back to at least to Hurwicz (1950) and it has been well documented in the literature. Theorem 3 requires  $N/T \rightarrow \kappa_1$  for the derivation of the asymptotic distribution of



$\hat{\pi}_{MG}$  due to the time series bias, and it is therefore unsuitable for panels with  $T$  small relative to  $N$ .

## 4 Bias-corrected CCEMG estimators

In this section we review the different procedures proposed in the literature for correcting the small sample time series bias of estimators in dynamic panels and consider the possibility of developing bias-corrected versions of CCEMG estimators for dynamic panels.

Existing literature focuses predominantly on homogeneous panels, where several different ways to correct for  $O(T^{-1})$  time series bias have been proposed. This literature can be divided into the following broad categories: (i) analytical corrections based on an asymptotic bias formula (Bruno, 2005, Bun, 2003, Bun and Carree, 2005 and 2006, Bun and Kiviet, 2003, Hahn and Kuersteiner, 2002 and 2011, Hahn and Moon, 2006, Hahn and Newey, 2004, Kiviet, 1995 and 1999, and Newey and Smith, 2004); (ii) bootstrap and simulation based bias corrections (Everaert and Ponzi, 2007, Phillips and Sul, 2003 and 2007), and (iii) other methods, including jackknife bias corrections (Hahn and Newey, 2004, and Dhaene and Jochmans, 2012) and the recursive mean adjustment correction procedures (So and Shin, 1999).

In contrast, bias correction for dynamic panels with heterogeneous coefficients have been considered only in few studies. Hsiao, Pesaran, and Tahmiscioglu (1999) investigate bias-corrected mean group estimation, where Kiviet and Phillips (1993) bias correction is applied to the individual estimates of short-run coefficients. Hsiao, Pesaran, and Tahmiscioglu (1999) propose also a Hierarchical Bayesian estimation of short-run coefficients, which they find to have good small sample properties in their Monte Carlo study.<sup>8</sup> Pesaran and Zhao (1999) investigate bias correction methods in estimating long-run coefficients and consider, in particular, two analytical corrections based on an approximation of the asymptotic bias of long-run coefficients, a bootstrap bias-corrected estimator, and a "naive" bias-corrected panel estimator computed from bias-corrected short-run coefficients (using a result derived by Kiviet and Phillips, 1993).

---

<sup>8</sup>Zhang and Small (2006) further develops the hierarchical Bayesian approach of Hsiao, Pesaran, and Tahmiscioglu (1999) by imposing a stationarity constraint on each of the cross section units and by considering different possibilities for starting values. Bayesian approach has also been developed by Canova and Marcet (1999) to study income convergence in a dynamic heterogeneous panel of countries, and by Canova and Ciccarelli (2004 and 2009) to forecast variables and turning points in a panel VAR. Forecasting with Bayesian shrinkage estimators have also been considered by Garcia-Ferrer, Highfield, Palm, and Zellner (1987), Zellner and Hong (1989) and Zellner, Hong, and Ki Min (1991).

## 4.1 Bias corrected versions of $\hat{\pi}_{MG}$

All the bias correction procedures reviewed above are developed for panel data models without unobserved common factors, and are not directly applicable to  $\hat{\pi}_{MG}$ . This applies to bootstrapped based corrections, as well as the analytical corrections based on asymptotic bias formulae such as the one derived by Kiviet and Phillips (1993). The development of analytical or bootstrapped bias correction procedures for dynamic panel data models with a multifactor error structure is beyond the scope of the present paper and deserve separate investigations of their own. Instead here we consider the application of jackknife and recursive mean adjustment bias correction procedures to  $\hat{\pi}_{MG}$  that do not require any knowledge of the error factor structure and are particularly simple to implement.

### 4.1.1 Jackknife bias correction

Jackknife bias correction is popular due to its simplicity and wide applicability. Jackknife bias correction can be applied to the panel mean group estimator, or at the level of unit-specific estimates. Since the mean group estimator is a linear function of the unit-specific estimators, applying the correction to  $\hat{\pi}_{MG}$  or to the unit-specific estimates,  $\hat{\pi}_i$ , yields numerically identical results. We consider the "half-panel jackknife" method discussed by Dhaene and Jochmans (2012), which corrects for  $O(T^{-1})$  bias. Jackknife bias-corrected CCEMG estimators are constructed as:

$$\tilde{\pi}_{MG} = 2\hat{\pi}_{MG} - \frac{1}{2} \left( \hat{\pi}_{MG}^a + \hat{\pi}_{MG}^b \right),$$

where  $\hat{\pi}_{MG}^a$  denotes the CCEMG estimator computed from the first half of the available time period, namely over the period  $t = 1, 2, \dots, [T/2]$ , where  $[T/2]$  denotes the integer part of  $T/2$ , and  $\hat{\pi}_{MG}^b$  is the CCEMG estimators computed using the observations over the period  $t = [T/2] + 1, [T/2] + 2, \dots, T$ .

### 4.1.2 Recursive mean adjustment

The second bias-correction is based on the recursive mean adjustment method proposed by So and Shin (1999), who advocate demeaning variables using the partial mean based on observations up

to the time period  $t - 1$ . In particular, we let

$$\tilde{y}_{it} = y_{it} - \frac{1}{t-1} \sum_{s=1}^{t-1} y_{is},$$

and

$$\tilde{\omega}_{it} = \omega_{it} - \frac{1}{t-1} \sum_{s=1}^{t-1} \omega_{is},$$

for  $i = 1, 2, \dots, N$  and  $t = 2, 3, \dots, T$ , where  $\omega_{it} = (\mathbf{x}'_{it}, \mathbf{g}'_{it})'$ . We then compute bias-adjusted CCE mean group estimator based on the recursive demeaned variables  $\tilde{y}_{it}$  and  $\tilde{\omega}_{it}$  (with  $T - 1$  available time periods,  $t = 2, 3, \dots, T$ ).

## 5 Monte Carlo Experiments

Our main objective is to investigate the small sample properties of the CCEMG estimator and its bias corrected versions in panel ARDL models under different assumptions concerning the parameter values and the degree of cross-sectional dependence. We also examine the robustness of the quasi maximum likelihood estimator (QMLE) developed by Moon and Weidner (2010a and 2010b) and the interactive-effects estimator (IFE) proposed by Bai (2009) to coefficients heterogeneity, and include an alternative MG estimator based on Song's extension of Bai's IFE approach (denoted as  $\hat{\pi}_{MG}^s$ ) and investigate its performance as well.

We start with the description of the data generating process in subsection 5.1, followed by a summary account of the different estimators being considered in subsection 5.2, before providing a summary of our main findings in the final subsection.

### 5.1 Data Generating Process

We set  $k_x = k_g = 1$  and write (1)-(3) as

$$y_{it} = c_{yi} + \phi_i y_{i,t-1} + \beta_{0i} x_{it} + \beta_{1i} x_{i,t-1} + u_{it}, \quad u_{it} = \gamma'_i \mathbf{f}_t + \varepsilon_{it}, \quad (33)$$

and

$$\begin{pmatrix} x_{it} \\ g_{it} \end{pmatrix} = \begin{pmatrix} c_{xi} \\ c_{gi} \end{pmatrix} + \begin{pmatrix} \alpha_{xi} \\ \alpha_{gi} \end{pmatrix} y_{i,t-1} + \begin{pmatrix} \gamma'_{xi} \\ \gamma'_{gi} \end{pmatrix} \mathbf{f}_t + \begin{pmatrix} v_{xit} \\ v_{git} \end{pmatrix}. \quad (34)$$

The unobserved common factors in  $\mathbf{f}_t$  and the unit-specific components  $\mathbf{v}_{it} = (v_{xit}, v_{git})'$  are generated as independent stationary AR(1) processes:

$$f_{t\ell} = \rho_{f\ell} f_{t-1,\ell} + \varsigma_{f t\ell}, \varsigma_{f t\ell} \sim IIDN(0, 1 - \rho_{f\ell}^2), \quad (35)$$

$$v_{xit} = \rho_{xi} v_{xi,t-1} + \varsigma_{xit}, \varsigma_{xit} \sim IIDN(0, \sigma_{vxi}^2), \quad (36)$$

$$v_{git} = \rho_{gi} v_{gi,t-1} + \varsigma_{git}, \varsigma_{git} \sim IIDN(0, \sigma_{vgi}^2) \quad (37)$$

for  $i = 1, 2, \dots, N$ ,  $\ell = 1, 2, \dots, m$ , and for  $t = -99, \dots, 0, 1, 2, \dots, T$  with the starting values  $f_{\ell,-100} = 0$ , and  $v_{xi,-100} = v_{gi,-100} = 0$ . The first 100 time observations ( $t = -99, -48, \dots, 0$ ) are discarded. We generate  $\rho_{xi}$  and  $\rho_{gi}$ , for  $i = 1, 2, \dots, N$  as *IIDU* [0.0.95], and consider two values for  $\rho_{f\ell}$ , representing the case of serially uncorrelated factors,  $\rho_{f\ell} = 0$ , for  $\ell = 1, 2, \dots, m$ , and the case of the serially correlated factors  $\rho_{f\ell} = 0.6$ , for  $\ell = 1, 2, \dots, m$ . We set  $\sigma_{vxi}^2 = \sigma_{vgi}^2 = \sigma_{vi}^2$  and allow  $\sigma_{vi}$  to be correlated with  $\beta_{0i}$  and set  $\sigma_{vi} = \beta_{i0} \sqrt{1 - [E(\rho_{xi})]^2}$ .

As before, we let  $\mathbf{z}_{it} = (y_{it}, x_{it}, g_{it})'$ , and write the data generating process for  $\mathbf{z}_{it}$  more compactly as (see (6)),

$$\mathbf{z}_{it} = \mathbf{c}_{zi} + \mathbf{A}_i \mathbf{z}_{i,t-1} + \mathbf{A}_{0i}^{-1} \mathbf{C}_i \mathbf{f}_t + \mathbf{A}_{0i}^{-1} \mathbf{e}_{it}, \quad (38)$$

where  $\mathbf{c}_{zi} = (c_{yi} + \beta_{0i} c_{xi}, c_{xi}, c_{gi})'$ ,

$$\mathbf{A}_i = \begin{pmatrix} \phi_i + \beta_{0i} \alpha_{xi} & \beta_{1i} & 0 \\ \alpha_{xi} & 0 & 0 \\ \alpha_{gi} & 0 & 0 \end{pmatrix}, \mathbf{A}_{0i}^{-1} = \begin{pmatrix} 1 & \beta_{0i} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \mathbf{C}_i = (\gamma_i, \gamma_{xi}, \gamma_{gi})',$$

and  $\mathbf{e}_{it} = (\varepsilon_{it} + \beta_{0i} v_{xit}, v_{xit}, v_{git})'$  is a serially correlated error vector. We generate  $\mathbf{z}_{it}$  for  $i = 1, 2, \dots, N$ , and  $t = -99, \dots, 0, 1, 2, \dots, T$  based on (38) with the starting values  $\mathbf{z}_{i,-100} = 0$ , and the first 100 time observations ( $t = -99, -48, \dots, 0$ ) are discarded as burn-in replications. The fixed effects are generated as  $c_{iy} \sim IIDN(1, 1)$ ,  $c_{xi} = c_{yi} + \varsigma_{cxi}$ , and  $c_{gi} = c_{yi} + \varsigma_{cgi}$ , where  $\varsigma_{cxi}, \varsigma_{cgi} \sim IIDN(0, 1)$ , thus allowing for dependence between  $(x_{it}, g_{it})'$  and  $c_{yi}$ .

For each  $i$  the process  $\{\mathbf{z}_{it}\}$  is stationary if  $\mathbf{f}_t$  and  $\mathbf{e}_{it}$  are stationary and the eigenvalues of  $\mathbf{A}_i$  lie inside the unit circle. More specifically the parameter choices for  $|\lambda_1(\mathbf{A}_i)| < 1$  have to be such that

$$\frac{1}{2} \left| \phi_i + \alpha_{xi} \beta_{0i} \pm \sqrt{(\phi_i + \alpha_{xi} \beta_{0i})^2 + 4\beta_{1i} \alpha_{xi}} \right| < 1.$$

Suppose now that we only consider positive values of  $\phi_i$ ,  $\alpha_{xi}$  and  $\beta_{0i}$ , such that  $\phi_i + \alpha_{xi}\beta_{0i} < 2$ . Then it is easily seen that sufficient stationary conditions are

$$\begin{aligned}(\beta_{0i} + \beta_{1i})\alpha_{xi} &< 1 - \phi_i, \\(\beta_{1i} - \beta_{0i})\alpha_{xi} &< 1 + \phi_i.\end{aligned}$$

Accordingly, we set  $\beta_{1i} = -0.5$  for all  $i$ , and generate  $\beta_{0i}$  as  $IIDU(0.5, 1)$ . When  $\alpha_{xi} > 0$ , we need to generate  $\alpha_{xi}$  such that  $0.5\alpha_{xi} < 1 - \phi_i$ . We consider two possibilities for  $\phi_i$ : Low values where  $\phi_i$  are generated as  $IIDU(0, 0.8)$  and  $\alpha_{xi}$  as  $IIDU(0, 0.35)$ . High values where use the draws,  $\phi_i \sim IIDU(0.5, 0.9)$  and  $\alpha_{xi} \sim IIDU(0, 0.15)$ . These choices ensure that the support of  $\lambda_1(\mathbf{A}_i)$  lies strictly inside the unit circle, as required by Assumption 5. Values of  $\alpha_{gi}$  do not affect the eigenvalues of  $\mathbf{A}_i$  and are generated as  $\alpha_{gi} \sim IIDU(0, 1)$ .

The above DGP is more general than the other DGPs used in other MC experiments in the literature and allows for weakly exogenous regressors. The factors and regressors are allowed to be correlated and persistent, and correlated fixed effects are included.

All factor loadings are generated independently as

$$\begin{aligned}\gamma_{i\ell} &= \gamma_\ell + \eta_{i,\gamma\ell}, \eta_{i,\gamma\ell} \sim IIDN(0, \sigma_{\gamma\ell}^2), \\ \gamma_{xil} &= \gamma_{x\ell} + \eta_{i,\gamma x\ell}, \eta_{i,\gamma x\ell} \sim IIDN(0, \sigma_{\gamma x\ell}^2), \\ \gamma_{gil} &= \gamma_{g\ell} + \eta_{i,\gamma g\ell}, \eta_{i,\gamma g\ell} \sim IIDN(0, \sigma_{\gamma g\ell}^2)\end{aligned}$$

for  $\ell = 1, 2, \dots, m$ , and  $i = 1, 2, \dots, N$ . Also, without loss of generality, the factor loadings are calibrated so that  $Var(\gamma'_i \mathbf{f}_t) = Var(\gamma'_{xi} \mathbf{f}_t) = Var(\gamma'_{gi} \mathbf{f}_t) = 1$ . We also set  $\sigma_{\gamma\ell}^2 = \sigma_{\gamma x\ell}^2 = \sigma_{\gamma g\ell}^2 = 0.2^2$ ,  $\gamma_\ell = \sqrt{b_{\gamma\ell}}$ ,  $\gamma_{x\ell} = \sqrt{\ell b_{x\ell}}$  and  $\gamma_{g\ell} = \sqrt{(2\ell - 1)b_{g\ell}}$ , for  $\ell = 1, 2, \dots, m$ , where  $b_\gamma = 1/m - \sigma_{\gamma\ell}^2$ ,  $b_x = 2/[m(m+1)] - 2/(m+1)\sigma_{x\ell}^2$  and  $b_g = 1/m^2 - \sigma_{g\ell}^2/m$ , for  $\ell = 1, 2, \dots, m$ . This ensures that the contribution of the unobserved factors to the variance of  $y_{it}$  does not rise with  $m$ . We consider  $m = 1, 2$  or 3 unobserved common factors.

Finally, the idiosyncratic errors,  $\varepsilon_{it}$ , are generated to be heteroskedastic and weakly cross-sectionally dependent. Specifically, we adopt the following spatial autoregressive model (SAR) to generate  $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, \dots, \varepsilon_{Nt})'$ :

$$\boldsymbol{\varepsilon}_t = a_\varepsilon \mathbf{S}_\varepsilon \boldsymbol{\varepsilon}_t + \mathbf{e}_{\varepsilon t}, \quad (39)$$

where the elements of  $\mathbf{e}_{\varepsilon t}$  are drawn as  $IIDN(0, \frac{1}{2}\sigma_i^2)$ , with  $\sigma_i^2$  obtained as independent draws from  $\chi^2(2)$  distribution,

$$\mathbf{S}_{\varepsilon} = \begin{pmatrix} 0 & \frac{1}{2} & 0 & 0 & \cdots & 0 \\ \frac{1}{2} & 0 & 1 & 0 & & 0 \\ 0 & 1 & 0 & \ddots & & \vdots \\ 0 & 0 & \ddots & \ddots & 1 & 0 \\ \vdots & & & & 1 & 0 & \frac{1}{2} \\ 0 & 0 & \cdots & 0 & \frac{1}{2} & 0 \end{pmatrix},$$

and the spatial autoregressive parameter is set to  $a_{\varepsilon} = 0.4$ . Note that  $\{\varepsilon_{it}\}$  is cross-sectionally weakly dependent for  $|a_{\varepsilon}| < 0.5$ .

In addition to these experiments, we also consider pure panel autoregressive experiments where we set  $\beta_{0i} = \beta_{1i} = 0$ , for all  $i$ . Table 1 summarizes the various parameter configurations of all the different experiments. In total, we conducted 24 experiments covering the various cases: with or without regressors in the equation for the dependent variable, low or high values of  $\phi = E(\phi_i)$ ,  $m = 1, 2$ , or 3 common factors, and persistent or serially uncorrelated common factors. We consider the following combinations of sample sizes:  $N, T \in \{40, 50, 100, 150, 200\}$ , and set the number of replications to  $R = 2000$ , in the case of all experiments.

## 5.2 Estimation techniques

The focus of the MC results will be on the estimates of the average parameter values  $\phi = E(\phi_i)$  and  $\beta_0 = E(\beta_{0i})$ , in the case of experiments with regressors,  $x_{it}$ . But before presenting the outcomes we briefly describe the computation of the alternative estimators being considered.<sup>9</sup>

### 5.2.1 Dynamic CCE mean group estimator

We base the CCE mean group estimator on the following cross-sectionally augmented unit-specific regressions,

$$y_{it} = c_{iy} + \phi_i y_{i,t-1} + \beta_{0i} x_{it} + \beta_{1i} x_{i,t-1} + \sum_{\ell=0}^{p_T} \delta'_{i\ell} \bar{\mathbf{z}}_{t-\ell} + e_{yit}, \quad (40)$$

---

<sup>9</sup>We are grateful to Jushan Bai, Hyungsik Roger Moon, and Martin Weidner for providing us with their Matlab codes.

for  $i = 1, 2, \dots, N$ , where  $\bar{\mathbf{z}}_t = N^{-1} \sum_{i=1}^N \mathbf{z}_{it} = (\bar{y}_t, \bar{x}_t, \bar{g}_t)'$ . We set  $p_T$  equal to the integer part of  $T^{1/3}$ , denoted as  $p_T = \lceil T^{1/3} \rceil$ . This gives the values of  $p_T = 3, 3, 4, 5, 5$  for  $T = 40, 50, 100, 150, 200$ , respectively. The CCE mean group estimator of  $\phi$  and  $\beta_0$  is then obtained by arithmetic averages of the least squares estimates of  $\phi_i$  and  $\beta_{0i}$  based on (40).

We also computed bias-corrected versions of the CCEMG estimator using the half-panel jack-knife and the recursive mean adjusted estimators as described in Section 4.1.

### 5.2.2 QMLE estimator by Moon and Weidner

We deal with fixed effects by de-meaning the variables before implementing the QMLE estimation procedure. Denote the demeaned variables as

$$\dot{y}_{it} = y_{it} - T^{-1} \sum_{t=1}^T y_{it}, \text{ and } \dot{x}_{it} = x_{it} - T^{-1} \sum_{t=1}^T x_{it}, \quad (41)$$

for  $s = 1, 2$  and  $i = 1, 2, \dots, N$ . We compute the bias-corrected QMLE estimator defined in Corollary 3.7 in Moon and Weidner (2010a) using  $\dot{y}_{it}$  as the dependent variable and the vector  $\dot{z}_{it} = (\dot{y}_{i,t-1}, \dot{x}_{it}, \dot{x}_{i,t-1})'$  as the vector of explanatory variables. Two options for the number of unobserved factors are considered: the true number of factors and the maximum number, 3, of unobserved factors.

### 5.2.3 Interactive-effects estimator by Bai

We deal with the fixed effects in the same way as before. In particular, we use the demeaned variables  $\dot{y}_{it}$ , and  $\dot{x}_{it,s}$  for  $s = 1, 2$ , to compute the interactive-effects estimator as the solution to the following set of non-linear equations:

$$\hat{\pi}_b = \left( \sum_{i=1}^N \dot{\Xi}_i' \mathbf{M}_{\hat{\mathbf{F}}} \dot{\Xi}_i \right)^{-1} \sum_{i=1}^N \dot{\Xi}_i' \mathbf{M}_{\hat{\mathbf{F}}} \dot{\mathbf{y}}_i, \quad (42)$$

$$\frac{1}{NT} \sum_{i=1}^N \left( \dot{\mathbf{y}}_i - \dot{\Xi}_i \hat{\pi}_b \right) \left( \dot{\mathbf{y}}_i - \dot{\Xi}_i \hat{\pi}_b \right)' \hat{\mathbf{F}} = \hat{\mathbf{F}} \hat{\mathbf{V}}, \quad (43)$$

where  $\hat{\pi}_b = (\hat{\phi}_b, \hat{\beta}_{0b}, \hat{\beta}_{1b})'$  is the interactive-effects estimator,  $\mathbf{M}_{\hat{\mathbf{F}}} = \mathbf{I}_T - \hat{\mathbf{F}} (\hat{\mathbf{F}} \hat{\mathbf{F}}')^{-1} \hat{\mathbf{F}}'$ ,  $\hat{\mathbf{V}}$  is a diagonal matrix with the  $m$  largest eigenvalues of the matrix  $\frac{1}{NT} \sum_{i=1}^N \left( \dot{\mathbf{y}}_i - \dot{\Xi}_i \hat{\pi}_b \right) \left( \dot{\mathbf{y}}_i - \dot{\Xi}_i \hat{\pi}_b \right)'$

arranged in decreasing order,  $\dot{\mathbf{y}}_i = (\dot{y}_{i2}, \dot{y}_{i3}, \dots, \dot{y}_{iT})'$  and

$$\dot{\mathbf{\Xi}}_i = \begin{pmatrix} \dot{y}_{i1} & \dot{x}_{i2} & \dot{x}_{i1} \\ \dot{y}_{i,2} & \dot{x}_{i3} & \dot{x}_{i2} \\ \vdots & \vdots & \vdots \\ \dot{y}_{i,T-1} & \dot{x}_{iT} & \dot{x}_{i,T-1} \end{pmatrix}.$$

The system of equations (42)-(43) is solved by an iterative method.

Bai (2009) does not allow for a lagged dependent variable in the derivation of the asymptotic results for the interactive-effects estimator, but considers this possibility in Monte Carlo experiments and concludes that parameters are well estimated also for the DGP with a lagged dependent variable. As in the case of the QMLE estimator, we consider Bai's estimates based on the true number of factors, and on the maximum number of factors, namely 3.

#### 5.2.4 Mean Group estimator based on Song's extension of Bai's IFE approach

Song (2013) extends Bai's IFE approach by allowing for coefficient heterogeneity and lags of the dependent variable. Song focuses on the estimates of individual coefficients obtained from the solution to the following system of nonlinear equations, which as he shows minimizes the sum of squared errors,

$$\hat{\boldsymbol{\pi}}_i^s = \left( \dot{\mathbf{\Xi}}_i' \mathbf{M}_{\hat{\mathbf{F}}} \dot{\mathbf{\Xi}}_i \right)^{-1} \dot{\mathbf{\Xi}}_i' \mathbf{M}_{\hat{\mathbf{F}}} \dot{\mathbf{y}}_i, \text{ for } i = 1, 2, \dots, N, \quad (44)$$

$$\frac{1}{NT} \sum_{i=1}^N \left( \dot{\mathbf{y}}_i - \dot{\mathbf{\Xi}}_i \hat{\boldsymbol{\pi}}_i \right) \left( \dot{\mathbf{y}}_i - \dot{\mathbf{\Xi}}_i \hat{\boldsymbol{\pi}}_i \right)' \hat{\mathbf{F}} = \hat{\mathbf{F}} \hat{\mathbf{V}}. \quad (45)$$

Similarly to Bai's IFE procedure, we use demeaned observations to deal with the presence of fixed effects and the system of equations (44)-(45) is solved numerically by an iterative method. Song (2013) establishes  $\sqrt{T}$  consistency rates of individual estimates  $\hat{\boldsymbol{\pi}}_i^s$  under asymptotics  $N, T \xrightarrow{j} \infty$  such that  $T/N^2 \rightarrow 0$ .

Given our random coefficient assumption on  $\boldsymbol{\pi}_i$ , we adopt the following mean group estimator based on Song's individual estimates,

$$\hat{\boldsymbol{\pi}}_{MG}^s = \frac{1}{N} \sum_{i=1}^N \hat{\boldsymbol{\pi}}_i^s,$$



and investigate the performance of  $\hat{\pi}_{MG}^s$  with its variance estimated nonparametrically by

$$\hat{\Sigma}_{MG}^s = \frac{1}{N-1} \sum_{i=1}^N (\hat{\pi}_i^s - \hat{\pi}_{MG}^s) (\hat{\pi}_i^s - \hat{\pi}_{MG}^s)'$$

Note that since  $\sqrt{T}(\hat{\pi}_i^s - \pi_i) = O_p(1)$  (uniformly in  $i$ ) as  $N, T \xrightarrow{j} \infty$  such that  $T/N^2 \rightarrow 0$  (see Song, 2013, Theorem 2), it readily follows that (also see Assumption 4)

$$\hat{\pi}_{MG}^s - \pi = \frac{1}{N} \sum_{i=1}^N v_{\pi i} + O_p\left(\frac{1}{\sqrt{T}}\right).$$

However, sufficient conditions for  $\sqrt{N}(\hat{\pi}_{MG}^s - \pi) \xrightarrow{d} N(0, \Omega_\pi)$  as  $N, T \xrightarrow{j} \infty$  remains to be investigated and this is outside the scope of the present paper.

## 6 Monte Carlo findings

In this section we report some of the main findings, and direct the reader to an online Supplement where the full set of results can be accessed.

Table 2 summarizes the results for the bias ( $\times 100$ ) and root mean square error (RMSE,  $\times 100$ ) in the case of the experiment with regressors,  $\phi = E(\phi_i) = 0.4$ , and one serially correlated unobserved common factor (Experiment 14 in Table 1). The first panel of this table gives the results for the fixed effects estimator (FE) which provides a benchmark against three sources of estimation bias: the time series bias of order  $T^{-1}$ , the bias from ignoring a serially correlated factor, and the bias due to coefficient (slope) heterogeneity. The latter two biases are not diminishing in  $T$  and we see that their combined effect remains substantial even for  $T = 200$ .

Next consider the QMLE estimator due to Moon and Weidner, which allows for unobserved factors, but fails to account for coefficient heterogeneity. As can be seen, this estimator still suffers from a substantial degree of heterogeneity bias which does not diminish in  $T$ . This is in line with the theoretical results derived in Pesaran and Smith (1995), where it is shown that in the presence of slope heterogeneity pooled least squares estimators are inconsistent in the case of panel data models with lagged dependent variables. This would have been the case even if the unobserved factors could have been estimated without any sampling errors. Initially, for  $T = 40$ , negative time series bias helps the performance of QMLE in our design, but as  $T$  increases, the time series bias

diminishes and the positive coefficient heterogeneity bias dominates the outcomes. The bias for  $T = 200$  ranges between 0.07 to 0.10 which amounts to 20 – 25% of the true value. Inclusion of 3 as opposed to 1 unobserved common factor improves the performance but does not mitigate fully the consequences of coefficient heterogeneity. Results for Bai’s IFE approach are similar to those of QMLE and are therefore reported only in the online Supplement to save space.

In contrast the CCEMG estimator deals with the presence of persistent factors and coefficient heterogeneity, but fails to adequately take account of the time series bias. As can be seen from the results, the uncorrected CCEMG estimator suffers from the time series bias when  $T$  is small, with the bias diminishing as  $T$  is increased. The sign of the bias is negative, which is in line with the existing literature. The bias of the CCEMG estimator is around  $-0.12$  for  $T = 40$ , and declines to around  $-0.02$  when  $T = 200$ .

Both bias correction methods considered are effective in reducing the time series bias of the CCEMG estimator, but the jackknife bias correction method turns out to be more successful overall. It is also interesting that the jackknife correction tends to slightly over-correct whereas the RMA procedure tends to under-correct. Both bias-correction methods also reduced the overall RMSE for all values of  $N$  and  $T$  considered.

The mean group estimator based on Song’s individual estimates performs slightly worse than the jackknife bias-corrected CCEMG, but overall its performance (in terms of bias and RMSE) seems to be satisfactory. The knowledge of the true number of factors, however, plays a very important role in improving the performance of this estimator.

Table 3 reports findings for estimation of  $\beta_0$  in the same experiment. As before, the FE and QMLE estimators continue to be biased even when  $T$  is large. The selection of the number factors seems to be quite important for the bias of QMLE estimator (and also Bai’s IFE estimator reported in the Supplement). The bias of CCEMG estimators is, in contrast, very small, between 0.0 to 0.02 for all values of  $N$  and  $T$ . Bias correction does not seem to matter for the CCEMG estimation of  $\beta_0$ . The small sample time series  $O(T^{-1})$  bias for the estimation of  $\beta_0$  is much smaller as compared to the bias of the autoregressive coefficient. Bias correction seems therefore not so important for the estimation of  $\beta_0$ , and the uncorrected version of CCEMG estimator performs better in terms of RMSE compared to its bias corrected versions.  $\hat{\pi}_{MG}^s$  also performs well although its RMSE is, in the majority of cases, slightly worse than RMSE of the uncorrected CCEMG estimator.

An important question is how robust are the various estimators to the number of unobserved factors. The MC results with more than one factor are summarized in Tables 4-7, and show that the CCEMG estimator continues to work well regardless of the number of factors and whether the factors are serially correlated. For  $m = 2$  or  $3$ , the performance of the CCEMG estimator and its bias-corrected versions is qualitatively similar to the case of  $m = 1$  discussed above. Only a slight deterioration in bias and RMSE is observed when  $m$  is increased to  $3$ , most likely due to the increased complexity encountered in approximating the space spanned by the unobserved common factors.

To check the validity of the asymptotic distribution of the CCEMG and other estimators, we now consider the size and power performance of the different estimators under consideration. We compute the size ( $\times 100$ ) at 5% nominal level and the power ( $\times 100$ ) for the estimation of  $\phi$  and  $\beta_0$  with the alternatives  $H_1 : \phi = 0.5$  and  $H_1 : \phi = 0.8$ , associated with the null values of  $\phi = 0.4$  and  $0.7$ , respectively, and the alternative of  $H_1 : \beta_0 = 0.85$ , associated with the null value of  $\beta_0 = 0.75$ . The results for size and power in the case of the Experiments 14, 16 and 18 are summarized in Tables 8-13.

As can be seen the tests based on FE and QMLE estimators and Bai's IFE (reported in the Supplement) are grossly oversized irrespective of whether the parameter of interest is  $\phi$  or  $\beta_0$ . In contrast the CCEMG estimator and the MG estimator based on Song's individual estimates have the correct size if one is interested in making inference about  $\beta_0$ , but both estimators tend to be over-sized if the aim is to make inference about  $\phi$ . These results are in line with our theoretical findings and largely reflect the time series bias of order  $O(T^{-1})$  which is present in the MG type estimators of  $\phi$ . The bias-corrected versions of the CCEMG estimator perform much better, with the jackknife bias-correction method generally outperforming the RMA procedure. The condition  $N/T \rightarrow \varkappa_1$ ,  $0 < \varkappa_1 < \infty$ , in Theorem 3 plays an important role in ensuring that the tests based on the CCEMG estimator of  $\phi$  have the correct size. In particular, the size worsens with an increase in the ratio  $N/T$ , especially when  $T = 40$ . Relatively good size (7%-9%) is achieved only when  $T > 100$ .

As already noted, the size of the tests based on the CCEMG estimator of  $\beta_0$ , (Tables 9, 11 and 12) is strikingly well behaved in all experiments and is very close to 5 percent for all values of  $N$  and  $T$ , which is in line with low biases reported for this estimator. Similar results also hold

for  $\pi_{MG}^s$ , although there are some incidences of size distortions for this MG estimator when  $T$  is relatively small (40 – 50).

Given the importance of the time series bias for the estimation of and inference on  $\phi$ , it is also reasonable to check the robustness of our findings to higher values of  $\phi$ . The estimation bias is likely to increase as  $\phi$  is increased towards unity. The results for the experiments with  $\phi$  set to 0.7 are reported in the online Supplement, and not surprisingly are generally worse than the results reported in the tables below for  $\phi = 0.4$ . Although, once again, the estimates of  $\beta_0$  tend not be much affected by the choice of  $\phi$ .

The results of the experiments with purely autoregressive panel data models (reported in the Supplement) are very similar to the ones discussed above, although the small sample performance of CCEMG estimator of  $\phi$  is slightly better as compared to the experiments with regressors.

Overall, our findings suggest that when  $\beta_0$  is the parameter of interest, the uncorrected CCEMG estimator seems to be preferred (in terms of bias, RMSE, size, and power), whereas jackknife corrected CCEMG estimator seems to be preferred for estimation of  $\phi$ , but the time dimension  $T$  needs to be relatively large in order to obtain a correct size for the tests of  $\phi$  based on the CCEMG type estimators of  $\phi$ , although some marginal improvements can be achieved if the jackknife bias-corrected version of CCEMG is used.

## 7 Conclusion

This paper extends the Common Correlated Effects (CCE) approach to estimation and inference in panel data models with a multi-factor error structure, originally proposed in Pesaran (2006), by allowing for the inclusion of lagged values of the dependent variable and weakly exogenous regressors in the panel data model. We show that the CCE mean group estimator continues to be valid asymptotically but the following two conditions must be satisfied to deal with the presence of lagged dependent variables amongst the regressors: a sufficient number of lags of cross-sectional averages must be included in individual equations, and the number of cross-sectional averages must be at least as large as the number of unobserved common factors. CCE mean group estimator and its jackknife and recursive mean adjustment bias corrected versions are easily implemented empirically. Results from an extensive set of Monte Carlo experiments show that the homogeneous slope estimators proposed in the literature can be seriously biased in the presence of

slope heterogeneity. In contrast the uncorrected CCEMG estimator proposed in the paper performs well (in terms of bias, RMSE, size and power) if the parameter of interest is the average slope of the regressors ( $\beta_0$ ), even if  $N$  and  $T$  are relatively small. But the situation is very different if the parameter of interest is the mean coefficient of the lagged dependent variable ( $\phi$ ). In the case of  $\phi$  the uncorrected CCEMG estimator suffers from the time series bias and tests based on it tend to be over-sized, unless  $T$  is sufficiently large relative to  $N$ . The jackknife bias-corrected CCEMG estimator, also proposed in the paper, does help in mitigating the time series bias, but it cannot fully deal with the size distortion unless  $T$  is sufficiently large. Improving on the small sample properties of the CCEMG estimators of  $\phi$  in the heterogeneous panel data models still remains a challenge to be taken on in the future.

**Table 1:** Parameters of the Monte Carlo Design

Exp.	Experiments without regressors ( $\beta_{0i} = \beta_{1i} = 0$ )			Exp.	Experiments with regressors ( $\beta_{0i} \sim IIDU [0.5, 1], \beta_{1i} = -0.5$ )		
	$\phi = E(\phi)$	$m$	$\rho_f$		$\phi = E(\phi)$	$m$	$\rho_f$
1	0.4	1	0	13	0.4	1	0
2	0.4	1	0.6	14	0.4	1	0.6
3	0.4	2	0	15	0.4	2	0
4	0.4	2	0.6	16	0.4	2	0.6
5	0.4	3	0	17	0.4	3	0
6	0.4	3	0.6	18	0.4	3	0.6
7	0.7	1	0	19	0.7	1	0
8	0.7	1	0.6	20	0.7	1	0.6
9	0.7	2	0	21	0.7	2	0
10	0.7	2	0.6	22	0.7	2	0.6
11	0.7	3	0	23	0.7	3	0
12	0.7	3	0.6	24	0.7	3	0.6

Notes: The dependent variable, regressors and covariates are generated according to (33)-(34) with  $\phi_i \sim IIDU [0, 0.8]$  (low value of  $\phi = E(\phi_i) = 0.4$ ) or with  $\phi_i \sim IIDU [0.5, 0.9]$  (high value of  $\phi = E(\phi_i) = 0.7$ ), with correlated fixed effects, and with cross-sectionally weakly dependent heteroskedastic idiosyncratic innovations generated from a SAR(1) model (39) with  $a_\varepsilon = 0.4$ . All experiments allow for feedback effects with  $\alpha_{xi} \sim IIDU [0, 0.35]$  for high value of  $\phi$ ,  $\alpha_{xi} \sim IIDU [0, 0.15]$  for low value of  $\phi$ , and  $\alpha_{gi} \sim IIDU [0, 1]$  for both values of  $\phi$ .

**Table 2.** Estimation of  $\phi$  in experiments with regressors,  $\phi = E(\phi_i) = 0.4$ , and  $m = 1$  correlated common factor. (Experiment 14)

(N,T)	Bias (x100)					RMSE (x100)				
	40	50	100	150	200	40	50	100	150	200
	<b>Fixed Effects estimates</b>									
40	13.12	14.74	17.83	18.80	19.61	15.48	16.72	19.12	19.83	20.55
50	13.08	14.79	18.07	19.25	19.60	15.13	16.50	19.14	20.12	20.41
100	13.42	15.11	18.29	19.53	20.12	15.08	16.43	19.00	20.12	20.64
150	13.95	15.05	18.47	19.67	20.23	15.47	16.20	19.09	20.09	20.61
200	13.47	15.27	18.64	19.71	20.23	14.89	16.38	19.21	20.11	20.57
	<b>Dynamic CCEMG without bias correction</b>									
40	-10.93	-8.25	-3.31	-1.98	-1.18	11.86	9.35	5.12	4.37	3.93
50	-11.12	-8.34	-3.61	-2.02	-1.30	11.88	9.23	5.02	4.05	3.74
100	-11.73	-9.04	-3.99	-2.41	-1.59	12.12	9.44	4.69	3.41	2.88
150	-12.06	-9.25	-4.22	-2.60	-1.76	12.33	9.54	4.68	3.25	2.62
200	-12.13	-9.37	-4.32	-2.68	-1.94	12.35	9.60	4.67	3.17	2.56
	<b>Dynamic CCEMG with RMA bias correction</b>									
40	-8.58	-5.82	-2.20	-0.84	-0.50	10.23	7.63	4.66	3.98	3.91
50	-8.55	-5.97	-2.14	-1.18	-0.57	9.92	7.47	4.24	3.77	3.44
100	-9.08	-6.17	-2.36	-1.25	-0.80	9.81	6.92	3.54	2.73	2.59
150	-9.29	-6.55	-2.40	-1.48	-0.89	9.80	7.06	3.24	2.49	2.22
200	-9.44	-6.75	-2.61	-1.47	-1.01	9.88	7.13	3.24	2.28	2.03
	<b>Dynamic CCEMG with jackknife bias correction</b>									
40	3.82	2.64	1.74	1.21	0.85	9.96	7.18	4.91	4.41	4.09
50	4.02	2.66	1.59	1.19	0.77	9.26	6.62	4.38	3.96	3.79
100	3.91	2.35	1.40	0.97	0.66	7.64	4.96	3.23	2.83	2.62
150	3.73	2.48	1.30	0.90	0.59	6.93	4.64	2.72	2.32	2.15
200	4.04	2.52	1.27	0.88	0.47	6.78	4.41	2.45	2.05	1.83
	<b>MG based on Song's individual estimates with 3 factors</b>									
40	-9.15	-6.77	-2.74	-1.38	-0.90	10.91	8.58	5.11	4.12	4.03
50	-9.48	-7.03	-2.76	-1.50	-0.95	10.81	8.38	4.52	3.84	3.54
100	-10.20	-7.32	-2.85	-1.72	-1.21	10.85	7.98	3.85	3.00	2.75
150	-10.53	-7.56	-2.98	-1.79	-1.27	10.99	8.02	3.69	2.74	2.33
200	-10.85	-7.78	-3.05	-1.85	-1.36	11.21	8.13	3.58	2.55	2.21
	<b>MG based on Song with true number of factors (<math>m = 1</math>)</b>									
40	-5.34	-3.95	-1.46	-0.40	-0.01	7.57	6.31	4.55	3.98	3.96
50	-6.03	-4.58	-1.76	-0.79	-0.28	7.61	6.33	4.06	3.60	3.43
100	-7.09	-5.47	-2.36	-1.40	-0.99	7.76	6.17	3.49	2.83	2.65
150	-7.27	-5.70	-2.56	-1.59	-1.11	7.71	6.17	3.33	2.60	2.24
200	-7.43	-5.87	-2.67	-1.67	-1.24	7.76	6.22	3.23	2.41	2.13
	<b>Moon and Weidner's QMLE with 3 factors</b>									
40	-2.67	0.94	5.73	7.30	7.73	8.93	7.99	8.68	9.55	9.82
50	-3.34	0.37	5.82	7.23	7.86	8.46	7.04	8.20	9.18	9.62
100	-4.66	-0.57	5.65	7.28	7.99	7.58	5.21	7.06	8.34	8.96
150	-5.74	-1.14	5.38	7.15	8.04	7.71	4.61	6.44	7.87	8.69
200	-6.05	-1.70	5.35	7.05	7.81	7.65	4.31	6.18	7.64	8.32
	<b>Moon and Weidner's QMLE with true number of factors (<math>m = 1</math>)</b>									
40	1.87	3.62	6.87	8.08	8.48	8.30	8.56	9.79	10.37	10.74
50	1.83	3.89	7.20	8.23	8.76	7.58	8.08	9.60	10.38	10.77
100	1.99	3.82	7.45	8.67	9.18	5.92	6.45	8.79	9.79	10.21
150	2.24	4.00	7.47	8.66	9.31	5.12	5.88	8.46	9.42	10.02
200	2.36	4.10	7.72	8.83	9.32	5.00	5.68	8.46	9.44	9.87

Notes: See notes to Table 1. CCEMG is based on (40) which features cross section averages of  $\mathbf{z}_{it} = (y_{it}, x_{it}, g_{it})'$ . QMLE estimator and MG estimator based on Song's individual estimates are computed from demeaned variables  $\hat{y}_{it}$  and  $\hat{x}_{it}$  defined in (41).

**Table 3.** Estimation of  $\beta_0$  in experiments with regressors,  $\phi = E(\phi_i) = 0.4$ , and  $m = 1$  correlated common factor. (Experiment 14)

(N,T)	Bias (x100)					RMSE (x100)				
	40	50	100	150	200	40	50	100	150	200
	<b>Fixed Effects estimates</b>									
40	51.52	51.31	51.66	51.37	51.49	51.96	51.68	51.88	51.54	51.64
50	50.96	51.08	51.27	51.25	51.33	51.37	51.42	51.47	51.40	51.46
100	51.07	51.13	51.36	51.13	51.35	51.40	51.39	51.52	51.24	51.43
150	51.22	51.11	51.25	51.22	51.32	51.54	51.36	51.38	51.32	51.39
200	50.99	51.28	51.20	51.09	51.20	51.27	51.51	51.32	51.17	51.27
	<b>Dynamic CCEMG without bias correction</b>									
40	1.37	1.14	0.69	0.45	0.18	5.92	5.28	3.70	3.30	3.08
50	1.05	0.82	0.48	0.28	0.27	5.48	4.59	3.37	2.93	2.84
100	1.11	0.92	0.58	0.30	0.23	3.92	3.37	2.45	2.15	1.93
150	1.23	1.05	0.46	0.26	0.28	3.34	2.88	1.98	1.77	1.61
200	1.24	0.97	0.50	0.33	0.26	2.97	2.51	1.77	1.52	1.37
	<b>Dynamic CCEMG with RMA bias correction</b>									
40	1.34	0.91	0.60	0.60	0.36	6.84	5.81	4.05	3.43	3.12
50	1.31	1.11	0.55	0.39	0.49	6.06	4.99	3.56	3.02	2.79
100	1.22	0.99	0.66	0.44	0.24	4.50	3.50	2.53	2.24	1.94
150	1.13	0.96	0.56	0.41	0.37	3.59	3.12	2.14	1.81	1.69
200	1.10	0.97	0.53	0.44	0.32	3.27	2.71	1.84	1.64	1.41
	<b>Dynamic CCEMG with jackknife bias correction</b>									
40	1.60	0.98	0.36	0.20	0.03	12.04	8.25	4.42	3.69	3.29
50	0.85	0.34	0.07	0.11	0.14	11.21	7.32	4.11	3.32	3.03
100	0.58	0.70	0.22	0.00	0.01	7.71	5.42	2.98	2.36	2.07
150	0.97	0.55	0.08	-0.06	0.07	6.49	4.32	2.38	1.99	1.71
200	0.84	0.52	0.08	0.03	0.02	5.65	3.88	2.08	1.68	1.44
	<b>MG based on Song's individual estimates with 3 factors</b>									
40	0.10	0.51	0.42	0.44	0.49	8.13	6.45	4.12	3.60	3.50
50	0.29	0.54	0.31	0.38	0.32	6.81	5.40	3.69	3.12	2.90
100	0.49	0.42	0.30	0.35	0.29	4.21	3.58	2.51	2.22	1.95
150	0.56	0.44	0.35	0.27	0.21	3.34	2.81	2.02	1.73	1.59
200	0.62	0.56	0.37	0.32	0.22	2.81	2.42	1.72	1.53	1.41
	<b>MG based on Song with true number of factors (<math>m = 1</math>)</b>									
40	-2.76	-2.08	-1.58	-1.51	-1.41	8.58	7.78	5.09	4.42	4.15
50	-1.67	-1.33	-1.09	-0.85	-0.95	7.50	5.61	4.09	3.36	3.25
100	0.09	0.04	-0.01	0.03	0.04	3.64	3.26	2.40	2.17	1.89
150	0.44	0.30	0.22	0.13	0.09	3.04	2.57	1.95	1.70	1.56
200	0.57	0.52	0.30	0.25	0.15	2.66	2.26	1.69	1.50	1.39
	<b>Moon and Weidner's QMLE with 3 factors</b>									
40	8.09	7.42	6.25	5.51	5.20	10.50	9.56	7.87	6.95	6.68
50	7.40	6.63	5.23	4.87	4.75	9.46	8.46	6.68	6.14	5.92
100	6.26	5.59	4.55	4.12	4.05	7.32	6.58	5.29	4.83	4.69
150	6.02	5.47	4.34	4.08	4.04	6.82	6.12	4.87	4.56	4.49
200	5.95	5.38	4.39	4.09	3.97	6.56	5.89	4.79	4.45	4.31
	<b>Moon and Weidner's QMLE with true number of factors (<math>m = 1</math>)</b>									
40	17.09	16.70	16.36	16.08	16.28	19.93	19.20	18.18	17.75	17.80
50	16.84	16.37	16.16	16.34	16.40	19.40	18.58	17.76	17.69	17.66
100	17.19	17.03	16.86	16.75	17.00	18.88	18.45	17.88	17.62	17.75
150	17.86	17.24	17.25	17.31	17.36	19.34	18.47	18.07	17.93	17.89
200	17.27	17.55	17.32	17.32	17.41	18.60	18.65	18.02	17.85	17.87

Notes: See notes to Table 1. CCEMG is based on (40) which features cross section averages of  $\mathbf{z}_{it} = (y_{it}, x_{it}, g_{it})'$ . QMLE estimator and MG estimator based on Song's individual estimates are computed from demeaned variables  $\hat{y}_{it}$  and  $\hat{x}_{it}$  defined in (41).



**Table 4.** Estimation of  $\phi$  in experiments with regressors,  $\phi = E(\phi_i) = 0.4$ , and  $m = 2$  correlated common factors. (Experiment 16)

(N,T)	Bias (x100)					RMSE (x100)				
	40	50	100	150	200	40	50	100	150	200
<b>Fixed Effects estimates</b>										
40	21.98	23.35	26.19	27.41	27.95	23.66	24.63	26.98	28.00	28.45
50	21.59	23.37	26.36	27.44	27.89	23.10	24.61	27.01	27.95	28.36
100	22.44	23.76	26.67	27.65	28.34	23.74	24.81	27.24	28.03	28.65
150	22.51	23.77	26.68	27.98	28.26	23.76	24.81	27.16	28.31	28.53
200	22.16	23.63	26.77	27.83	28.42	23.37	24.61	27.22	28.13	28.68
<b>Dynamic CCEMG without bias correction</b>										
40	-10.66	-7.93	-3.13	-1.58	-0.68	11.66	9.15	5.12	4.21	3.93
50	-10.83	-8.07	-3.23	-1.66	-0.87	11.64	9.02	4.82	3.84	3.64
100	-11.18	-8.31	-3.43	-1.94	-1.20	11.61	8.79	4.28	3.14	2.66
150	-11.45	-8.67	-3.67	-2.02	-1.37	11.74	8.99	4.23	2.87	2.40
200	-11.64	-8.87	-3.78	-2.23	-1.42	11.86	9.11	4.19	2.85	2.23
<b>Dynamic CCEMG with RMA bias correction</b>										
40	-8.72	-5.77	-1.98	-0.89	-0.14	10.40	7.66	4.65	4.08	3.89
50	-8.77	-5.88	-2.10	-0.97	-0.38	10.11	7.37	4.29	3.65	3.57
100	-9.14	-6.11	-2.30	-1.28	-0.75	9.89	6.94	3.51	2.83	2.53
150	-9.33	-6.42	-2.45	-1.33	-0.88	9.89	6.97	3.28	2.48	2.18
200	-9.49	-6.56	-2.53	-1.48	-0.87	9.92	7.00	3.17	2.33	1.95
<b>Dynamic CCEMG with jackknife bias correction</b>										
40	3.94	2.97	1.93	1.54	1.40	10.00	7.26	5.01	4.45	4.24
50	4.11	2.86	1.79	1.50	1.14	9.39	6.51	4.50	4.02	3.82
100	3.96	2.83	1.63	1.17	0.79	7.73	5.33	3.39	2.92	2.60
150	4.10	2.59	1.45	1.11	0.63	7.18	4.69	2.80	2.46	2.17
200	4.12	2.70	1.46	0.99	0.64	6.88	4.53	2.55	2.12	1.89
<b>MG based on Song's individual estimates with 3 factors</b>										
40	-9.08	-6.33	-2.04	-0.82	-0.32	10.77	8.02	4.56	4.11	3.94
50	-9.02	-6.41	-1.91	-0.94	-0.36	10.26	7.80	4.12	3.61	3.54
100	-9.46	-6.79	-2.29	-1.01	-0.61	10.10	7.49	3.48	2.69	2.56
150	-9.83	-6.89	-2.39	-1.25	-0.75	10.28	7.37	3.21	2.42	2.15
200	-10.30	-7.19	-2.61	-1.37	-0.85	10.64	7.54	3.21	2.24	1.97
<b>MG based on Song with true number of factors (<math>m = 2</math>)</b>										
40	-7.57	-5.41	-1.76	-0.62	-0.14	9.20	7.18	4.39	4.04	3.88
50	-7.54	-5.48	-1.62	-0.79	-0.22	8.80	6.90	3.97	3.57	3.52
100	-7.86	-5.87	-2.04	-0.85	-0.47	8.49	6.57	3.31	2.62	2.51
150	-8.13	-5.91	-2.12	-1.09	-0.61	8.55	6.41	3.00	2.35	2.09
200	-8.39	-6.08	-2.32	-1.19	-0.71	8.72	6.44	2.97	2.13	1.90
<b>Moon and Weidner's QMLE with 3 factors</b>										
40	-0.27	3.31	8.40	9.94	10.80	8.95	8.83	10.68	11.76	12.41
50	-1.40	2.26	7.69	9.31	9.96	8.47	7.59	9.65	10.86	11.44
100	-4.23	0.15	6.46	8.16	9.04	7.52	5.54	7.77	9.11	9.80
150	-5.76	-1.28	5.77	7.80	8.49	7.56	4.73	6.79	8.53	9.12
200	-6.44	-1.76	5.41	7.32	8.23	7.76	4.23	6.19	7.90	8.74
<b>Moon and Weidner's QMLE with true number of factors (<math>m = 2</math>)</b>										
40	2.89	5.33	9.61	10.97	11.66	8.99	9.32	11.73	12.80	13.26
50	2.09	4.49	8.85	10.26	10.79	8.15	8.42	10.77	11.77	12.27
100	0.23	3.14	7.60	8.96	9.77	5.46	5.82	8.70	9.83	10.50
150	-0.15	2.59	7.53	9.15	9.77	4.49	4.82	8.29	9.75	10.30
200	-0.37	2.64	7.56	9.13	9.85	3.91	4.39	8.14	9.59	10.28

Notes: See notes to Table 1. CCEMG is based on (40) which features cross section averages of  $\mathbf{z}_{it} = (y_{it}, x_{it}, g_{it})'$ . QMLE estimator and MG estimator based on Song's individual estimates are computed from demeaned variables  $\hat{y}_{it}$  and  $\hat{x}_{it}$  defined in (41).

**Table 5.** Estimation of  $\beta_0$  in experiments with regressors,  $\phi = E(\phi_i) = 0.4$ , and  $m = 2$  correlated common factors. (Experiment 16)

(N,T)	Bias (x100)					RMSE (x100)				
	40	50	100	150	200	40	50	100	150	200
	<b>Fixed Effects estimates</b>									
40	9.94	9.49	9.66	9.61	9.70	14.19	13.18	11.72	11.26	11.02
50	9.95	9.43	9.53	9.87	9.92	13.87	12.86	11.48	11.28	11.12
100	9.85	9.83	9.85	9.80	9.46	13.47	12.63	11.49	11.02	10.45
150	10.15	9.75	9.86	9.86	9.74	13.56	12.58	11.28	10.89	10.59
200	9.62	9.81	9.87	9.95	9.60	13.04	12.49	11.35	10.92	10.40
	<b>Dynamic CCEMG without bias correction</b>									
40	1.00	0.71	0.43	0.09	0.13	5.75	5.10	3.82	3.31	3.08
50	0.79	0.76	0.24	0.24	0.16	5.23	4.57	3.38	3.00	2.77
100	0.95	0.73	0.30	0.15	-0.01	3.78	3.32	2.40	2.10	1.93
150	1.06	0.61	0.28	0.23	0.07	3.26	2.75	1.98	1.78	1.58
200	0.98	0.75	0.29	0.17	0.08	2.80	2.34	1.71	1.48	1.37
	<b>Dynamic CCEMG with RMA bias correction</b>									
40	1.12	0.80	0.58	0.24	0.27	6.59	5.65	4.00	3.44	3.12
50	0.82	0.73	0.42	0.41	0.33	5.89	4.95	3.59	3.10	2.81
100	0.99	0.73	0.45	0.33	0.18	4.25	3.58	2.50	2.18	1.98
150	1.07	0.68	0.41	0.40	0.24	3.66	3.02	2.09	1.84	1.63
200	0.98	0.79	0.43	0.30	0.23	3.12	2.54	1.83	1.56	1.43
	<b>Dynamic CCEMG with jackknife bias correction</b>									
40	1.42	0.54	0.20	0.01	0.06	12.35	8.24	4.62	3.73	3.28
50	0.94	0.45	0.12	0.15	0.12	10.68	7.40	4.05	3.35	2.93
100	0.89	0.52	0.09	0.10	-0.03	7.61	5.17	2.89	2.40	2.09
150	1.22	0.44	0.10	0.11	0.03	6.44	4.42	2.42	1.97	1.70
200	0.95	0.67	0.08	0.01	0.03	5.72	3.73	2.10	1.68	1.49
	<b>MG based on Song's individual estimates with 3 factors</b>									
40	0.98	0.52	0.40	0.39	0.18	7.45	5.95	3.94	3.55	3.33
50	0.77	0.59	0.38	0.38	0.32	6.31	5.35	3.65	3.15	3.00
100	0.77	0.77	0.43	0.39	0.33	4.17	3.58	2.64	2.34	2.21
150	0.91	0.70	0.40	0.41	0.39	3.41	2.93	2.22	1.97	1.88
200	0.96	0.75	0.54	0.44	0.35	2.92	2.50	1.92	1.74	1.78
	<b>MG based on Song with true number of factors (<math>m = 2</math>)</b>									
40	0.87	0.69	0.54	0.43	0.27	6.71	5.58	3.89	3.50	3.25
50	0.82	0.67	0.35	0.43	0.34	5.68	4.96	3.52	3.10	2.92
100	0.90	0.84	0.51	0.40	0.41	3.88	3.43	2.54	2.26	2.16
150	0.94	0.78	0.45	0.43	0.43	3.27	2.88	2.12	1.90	1.79
200	1.00	0.77	0.58	0.45	0.34	2.83	2.41	1.86	1.68	1.70
	<b>Moon and Weidner's QMLE with 3 factors</b>									
40	5.21	4.83	4.53	4.20	4.23	7.88	7.45	6.43	5.81	5.89
50	5.06	4.95	4.47	4.57	4.49	7.55	7.08	6.04	5.94	5.78
100	5.54	5.14	4.81	4.53	4.47	6.83	6.29	5.66	5.27	5.18
150	5.62	5.15	4.66	4.57	4.43	6.54	5.95	5.25	5.11	4.91
200	5.68	5.21	4.56	4.45	4.31	6.36	5.81	5.04	4.84	4.69
	<b>Moon and Weidner's QMLE with true number of factors (<math>m = 2</math>)</b>									
40	4.94	4.68	4.32	3.95	4.05	7.94	7.52	6.41	5.77	5.86
50	4.91	4.83	4.33	4.43	4.33	7.62	7.08	6.04	5.96	5.74
100	5.43	5.18	4.91	4.65	4.64	6.89	6.47	5.82	5.45	5.39
150	5.59	5.33	4.98	4.96	4.83	6.68	6.27	5.64	5.51	5.34
200	5.75	5.40	5.02	5.00	4.90	6.64	6.16	5.55	5.43	5.31

Notes: See notes to Table 1. CCEMG is based on (40) which features cross section averages of  $\mathbf{z}_{it} = (y_{it}, x_{it}, g_{it})'$ . QMLE estimator and MG estimator based on Song's individual estimates are computed from demeaned variables  $\hat{y}_{it}$  and  $\hat{x}_{it}$  defined in (41).

**Table 6.** Estimation of  $\phi$  in experiments with regressors,  $\phi = E(\phi_i) = 0.4$ , and  $m = 3$  correlated common factors. (Experiment 18)

(N,T)	Bias (x100)					RMSE (x100)				
	40	50	100	150	200	40	50	100	150	200
	<b>Fixed Effects estimates</b>									
40	25.74	27.09	30.09	31.21	31.58	27.09	28.16	30.68	31.67	31.98
50	25.86	27.55	30.06	31.22	31.77	27.17	28.50	30.60	31.65	32.12
100	26.31	27.72	30.40	31.34	31.70	27.37	28.58	30.84	31.65	31.95
150	26.16	27.50	30.46	31.58	31.94	27.14	28.36	30.87	31.87	32.15
200	26.26	27.65	30.51	31.62	32.21	27.26	28.40	30.89	31.88	32.42
	<b>Dynamic CCEMG without bias correction</b>									
40	-11.29	-8.46	-3.08	-1.41	-0.61	12.26	9.53	5.06	4.12	4.04
50	-11.36	-8.38	-3.34	-1.61	-0.72	12.16	9.26	4.88	3.84	3.56
100	-11.59	-8.71	-3.50	-1.74	-1.09	12.00	9.14	4.32	3.02	2.66
150	-11.64	-8.76	-3.53	-1.88	-1.13	11.94	9.07	4.09	2.75	2.29
200	-11.64	-8.81	-3.62	-1.93	-1.13	11.86	9.05	4.03	2.61	2.07
	<b>Dynamic CCEMG with RMA bias correction</b>									
40	-9.99	-6.82	-2.32	-1.04	-0.42	11.45	8.41	4.78	4.04	4.04
50	-10.02	-6.86	-2.59	-1.32	-0.58	11.26	8.18	4.52	3.79	3.56
100	-10.44	-7.26	-2.84	-1.53	-1.03	11.13	7.94	3.88	2.97	2.66
150	-10.56	-7.34	-2.93	-1.72	-1.09	11.08	7.84	3.62	2.67	2.29
200	-10.56	-7.37	-3.03	-1.77	-1.15	10.95	7.77	3.56	2.51	2.09
	<b>Dynamic CCEMG with jackknife bias correction</b>									
40	4.25	2.99	2.17	1.78	1.37	10.26	7.47	5.08	4.47	4.34
50	4.49	3.12	1.90	1.56	1.21	9.65	6.91	4.52	4.05	3.79
100	3.74	2.77	1.71	1.30	0.73	7.59	5.35	3.36	2.96	2.61
150	3.99	2.78	1.58	1.10	0.67	7.19	4.92	2.83	2.41	2.15
200	4.24	2.60	1.50	1.05	0.62	6.99	4.49	2.54	2.14	1.90
	<b>MG based on Song with true number of factors (<math>m = 3</math>)</b>									
40	-7.94	-4.88	-0.14	0.96	1.54	9.72	6.96	4.17	3.95	4.10
50	-7.86	-5.05	-0.38	0.76	1.32	9.35	6.75	3.77	3.66	3.70
100	-8.79	-5.82	-0.95	0.28	0.73	9.58	6.65	2.83	2.51	2.67
150	-9.28	-6.28	-1.51	-0.30	0.19	9.78	6.84	2.69	2.18	2.03
200	-9.86	-6.76	-1.96	-0.70	-0.21	10.23	7.19	2.78	1.97	1.80
	<b>Moon and Weidner's QMLE with true number of factors (<math>m = 3</math>)</b>									
40	2.21	5.83	11.43	12.87	13.18	9.75	10.12	13.26	14.43	14.64
50	0.88	4.70	10.13	11.66	12.49	8.75	8.96	11.79	12.99	13.71
100	-3.20	0.99	7.93	9.91	10.64	7.18	5.48	9.02	10.72	11.33
150	-5.01	-0.42	6.91	9.05	9.88	7.07	4.54	7.75	9.65	10.39
200	-5.70	-1.20	6.25	8.49	9.54	7.01	4.00	6.94	8.97	9.97

Notes: See notes to Table 1. CCEMG is based on (40) which features cross section averages of  $\mathbf{z}_{it} = (y_{it}, x_{it}, g_{it})'$ . QMLE estimator and MG estimator based on Song's individual estimates are computed from demeaned variables  $\hat{y}_{it}$  and  $\hat{x}_{it}$  defined in (41).

**Table 7.** Estimation of  $\beta_0$  in experiments with regressors,  $\phi = E(\phi_i) = 0.4$ , and  $m = 3$  correlated common factors. (Experiment 18)

(N,T)	Bias (x100)					RMSE (x100)				
	40	50	100	150	200	40	50	100	150	200
	<b>Fixed Effects estimates</b>									
40	-18.62	-18.43	-18.70	-18.50	-18.38	21.16	20.51	19.99	19.45	19.25
50	-18.31	-18.45	-18.29	-18.80	-18.64	20.83	20.42	19.47	19.70	19.41
100	-18.20	-18.56	-18.40	-18.29	-18.42	20.42	20.29	19.32	18.98	18.98
150	-18.10	-18.24	-18.43	-18.45	-18.32	20.18	19.91	19.33	19.04	18.82
200	-17.87	-18.42	-18.44	-18.73	-18.54	19.90	20.08	19.23	19.31	18.99
	<b>Dynamic CCEMG without bias correction</b>									
40	0.98	0.84	0.47	0.38	0.27	6.12	5.22	3.76	3.31	3.12
50	0.93	0.73	0.53	0.34	0.05	5.30	4.67	3.43	2.96	2.71
100	0.92	0.66	0.32	0.18	0.10	3.78	3.39	2.42	2.07	1.94
150	0.83	0.65	0.40	0.12	0.15	3.23	2.76	1.94	1.72	1.61
200	0.90	0.73	0.29	0.13	0.11	2.81	2.47	1.67	1.49	1.36
	<b>Dynamic CCEMG with RMA bias correction</b>									
40	1.01	0.86	0.52	0.46	0.35	6.91	5.63	3.92	3.42	3.18
50	0.73	0.67	0.59	0.45	0.16	6.00	5.15	3.61	3.04	2.72
100	0.93	0.62	0.41	0.28	0.20	4.30	3.63	2.53	2.13	1.99
150	0.81	0.58	0.48	0.24	0.27	3.61	3.03	2.06	1.80	1.65
200	0.87	0.67	0.38	0.25	0.22	3.17	2.67	1.75	1.56	1.40
	<b>Dynamic CCEMG with jackknife bias correction</b>									
40	1.02	0.93	0.22	0.19	0.15	12.39	8.52	4.56	3.76	3.36
50	1.05	0.68	0.29	0.19	-0.06	10.94	7.73	4.21	3.34	2.91
100	1.39	0.45	0.10	-0.01	0.02	7.99	5.34	2.91	2.33	2.10
150	1.01	0.54	0.17	-0.03	0.09	6.52	4.44	2.32	1.95	1.72
200	1.00	0.58	0.03	-0.01	0.05	5.72	3.88	2.01	1.69	1.47
	<b>MG based on Song with true number of factors (<math>m = 3</math>)</b>									
40	0.49	0.24	-0.21	-0.08	0.01	7.73	6.23	4.20	3.77	3.59
50	0.20	0.29	0.02	-0.08	-0.09	6.71	5.55	3.91	3.34	3.12
100	0.38	0.26	-0.02	-0.30	-0.19	4.28	3.67	2.78	2.52	2.44
150	0.27	0.28	-0.12	-0.25	-0.20	3.29	2.88	2.32	2.12	2.10
200	0.35	0.22	-0.07	-0.22	-0.22	2.84	2.47	1.95	1.82	1.80
	<b>Moon and Weidner's QMLE with true number of factors (<math>m = 3</math>)</b>									
40	4.18	4.51	4.13	4.24	4.12	7.19	7.15	6.10	5.94	5.74
50	4.67	4.75	4.36	4.17	4.08	6.96	6.75	5.92	5.56	5.45
100	5.17	4.88	4.61	4.48	4.46	6.36	5.97	5.38	5.17	5.10
150	5.19	5.01	4.75	4.44	4.57	6.10	5.79	5.24	4.93	5.03
200	5.28	5.15	4.67	4.44	4.41	5.94	5.75	5.07	4.82	4.77

Notes: See notes to Table 1. CCEMG is based on (40) which features cross section averages of  $\mathbf{z}_{it} = (y_{it}, x_{it}, g_{it})'$ . QMLE estimator and MG estimator based on Song's individual estimates are computed from demeaned variables  $\hat{y}_{it}$  and  $\hat{x}_{it}$  defined in (41).

**Table 8.** Size and Power of estimating  $\phi$  in Experiment 14 (with regressors,  $\phi = 0.4$ ,  $m = 1$  and  $\rho_f = 0.6$ ).

(N,T)	Size (x100)					Power (x100)				
	40	50	100	150	200	40	50	100	150	200
	<b>Fixed Effects estimates</b>									
40	88.00	91.70	98.10	99.65	99.80	60.85	68.50	83.65	89.45	92.10
50	88.75	94.15	99.30	99.90	99.85	63.25	70.20	85.65	92.00	93.80
100	94.85	98.10	99.90	100.00	100.00	71.35	75.95	91.90	96.20	97.80
150	96.70	99.05	100.00	100.00	100.00	77.30	81.00	95.60	98.65	99.45
200	97.15	99.35	100.00	100.00	100.00	78.40	83.15	96.45	99.15	99.65
	<b>Dynamic CCEMG without bias correction</b>									
40	72.75	53.85	16.00	10.80	8.90	99.80	99.35	94.15	90.15	88.25
50	80.95	60.75	21.45	12.90	10.00	100.00	99.90	98.00	95.20	92.40
100	98.30	92.60	38.80	19.80	11.90	100.00	100.00	100.00	99.90	99.90
150	99.95	98.60	57.45	28.50	16.75	100.00	100.00	100.00	100.00	100.00
200	100.00	99.70	70.85	37.15	22.10	100.00	100.00	100.00	100.00	100.00
	<b>Dynamic CCEMG with RMA bias correction</b>									
40	42.65	27.85	10.15	6.20	6.90	94.90	92.40	86.25	80.70	78.80
50	49.65	32.65	10.55	8.25	6.35	97.65	96.65	91.90	89.25	89.05
100	79.50	57.80	17.20	8.30	7.20	100.00	100.00	99.85	99.70	99.55
150	91.80	77.40	22.70	11.75	9.20	100.00	100.00	100.00	100.00	100.00
200	95.60	88.95	32.40	14.55	10.55	100.00	100.00	100.00	100.00	100.00
	<b>Dynamic CCEMG with jackknife bias correction</b>									
40	14.15	12.20	9.85	8.65	7.90	20.65	30.10	49.60	59.90	65.00
50	15.40	12.60	9.05	8.15	8.95	21.20	33.70	59.80	69.20	75.00
100	21.70	16.05	10.80	7.80	7.65	34.05	54.95	86.80	93.80	96.25
150	26.85	20.35	10.95	8.65	7.05	42.20	66.00	96.55	99.15	99.50
200	31.90	25.85	11.95	8.60	6.95	48.15	74.15	99.00	99.85	99.95
	<b>MG based on Song's individual estimates with 3 factors</b>									
40	51.50	36.20	13.75	7.95	7.15	95.60	94.10	88.20	84.55	81.55
50	62.00	45.45	13.15	9.05	7.05	98.45	98.50	95.65	91.90	89.50
100	90.30	75.70	23.00	11.35	8.95	100.00	100.00	99.95	99.70	99.70
150	97.35	89.50	33.45	16.70	10.60	100.00	100.00	100.00	100.00	100.00
200	99.50	96.20	42.95	20.80	13.60	100.00	100.00	100.00	100.00	100.00
	<b>MG based on Song with true number of factors (<math>m = 1</math>)</b>									
40	30.45	20.90	10.35	6.55	7.45	91.45	89.00	81.55	76.70	74.85
50	39.85	27.70	10.50	7.15	6.75	96.45	95.85	91.70	88.70	86.00
100	72.45	56.80	17.60	10.45	8.30	100.00	100.00	99.90	99.60	99.55
150	88.60	74.85	26.75	15.10	9.15	100.00	100.00	100.00	100.00	100.00
200	95.45	87.60	34.80	17.10	11.70	99.95	100.00	100.00	100.00	100.00
	<b>Moon and Weidner's QMLE with 3 factors</b>									
40	51.95	52.55	71.60	80.20	84.90	81.85	74.10	66.30	67.80	70.55
50	55.15	51.30	74.35	83.35	87.85	85.05	79.20	67.45	69.60	72.35
100	63.80	50.15	81.35	91.85	94.90	96.50	91.50	73.20	70.80	73.05
150	73.00	53.45	84.25	96.10	98.55	99.10	96.55	80.40	74.40	70.95
200	79.65	57.00	89.15	97.60	99.05	99.55	98.80	84.85	77.95	75.15
	<b>Moon and Weidner's QMLE with true number of factors (<math>m = 1</math>)</b>									
40	46.30	53.35	72.50	81.15	85.55	67.95	63.15	63.80	65.95	68.25
50	46.15	53.50	76.90	83.45	88.70	69.80	67.05	65.35	68.05	72.60
100	49.60	57.30	88.10	94.60	96.45	80.85	76.50	66.55	67.90	70.05
150	50.65	64.35	93.10	98.50	98.95	86.30	80.85	70.95	68.85	70.60
200	54.20	70.75	96.85	99.30	99.55	88.40	84.10	69.30	68.80	71.75

Notes: See notes to Table 1. CCEMG is based on (40) which features cross section averages of  $\mathbf{z}_{it} = (y_{it}, x_{it}, g_{it})'$ . QMLE estimator and MG estimator based on Song's individual estimates are computed from demeaned variables  $\hat{y}_{it}$  and  $\hat{x}_{it}$  defined in (41).

**Table 9.** Size and Power of estimating  $\beta_0$  in Experiment 14 (with regressors,  $\phi = 0.4$ ,  $m = 1$  and  $\rho_f = 0.6$ ).

(N,T)	Size (x100)					Power (x100)				
	40	50	100	150	200	40	50	100	150	200
<b>Fixed Effects estimates</b>										
<b>40</b>	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
<b>50</b>	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
<b>100</b>	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
<b>150</b>	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
<b>200</b>	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
<b>Dynamic CCEMG without bias correction</b>										
<b>40</b>	6.75	7.15	6.85	7.10	6.60	33.35	41.45	74.80	85.20	91.10
<b>50</b>	7.00	6.25	5.05	5.95	7.20	42.55	54.15	82.70	92.45	94.60
<b>100</b>	6.95	6.45	6.10	5.65	5.10	67.05	80.30	98.10	99.65	100.00
<b>150</b>	7.90	7.40	6.30	6.10	5.85	81.55	92.60	99.85	99.95	100.00
<b>200</b>	8.65	7.45	7.00	5.80	5.65	90.80	97.30	99.95	100.00	100.00
<b>Dynamic CCEMG with RMA bias correction</b>										
<b>40</b>	6.35	7.65	7.25	6.70	6.45	28.05	40.55	69.55	81.90	90.00
<b>50</b>	6.55	6.25	6.40	6.45	6.55	33.45	44.95	78.75	90.40	94.40
<b>100</b>	7.40	5.80	6.80	7.15	4.70	56.40	74.60	97.45	99.65	100.00
<b>150</b>	7.10	7.95	6.35	6.85	6.50	74.45	88.10	99.80	100.00	100.00
<b>200</b>	8.05	7.25	6.30	7.20	5.50	84.15	94.70	100.00	100.00	100.00
<b>Dynamic CCEMG with jackknife bias correction</b>										
<b>40</b>	5.55	6.40	5.55	6.55	6.50	62.10	87.15	99.70	99.95	100.00
<b>50</b>	6.40	4.85	5.65	6.45	6.90	68.45	92.00	100.00	100.00	100.00
<b>100</b>	5.00	6.70	6.00	5.20	5.10	90.00	99.75	100.00	100.00	100.00
<b>150</b>	5.85	5.45	4.95	6.45	5.60	97.60	100.00	100.00	100.00	100.00
<b>200</b>	6.20	5.95	5.40	5.00	4.30	99.45	100.00	100.00	100.00	100.00
<b>MG based on Song's individual estimates with 3 factors</b>										
<b>40</b>	4.30	5.15	4.20	4.60	4.20	30.15	37.45	61.55	71.90	75.50
<b>50</b>	5.20	4.50	4.45	4.30	3.75	36.60	48.30	75.50	84.00	87.60
<b>100</b>	5.45	6.00	5.80	5.85	4.85	68.15	79.75	97.05	99.50	99.75
<b>150</b>	6.30	5.55	5.95	5.05	5.20	85.80	93.70	99.70	99.80	100.00
<b>200</b>	6.50	6.80	6.00	6.00	5.85	93.15	98.30	100.00	100.00	100.00
<b>MG based on Song with true number of factors (<math>m = 1</math>)</b>										
<b>40</b>	8.30	9.05	5.80	6.25	6.45	51.45	56.15	78.15	85.10	88.20
<b>50</b>	8.60	6.75	6.55	4.95	4.90	55.60	66.10	85.55	92.25	95.00
<b>100</b>	5.55	6.45	5.25	5.75	4.90	80.80	89.25	98.10	99.60	99.95
<b>150</b>	7.25	6.25	5.60	5.65	4.85	92.30	96.65	99.95	100.00	100.00
<b>200</b>	7.00	6.55	6.15	6.15	5.70	97.20	99.35	99.95	100.00	100.00
<b>Moon and Weidner's QMLE with 3 factors</b>										
<b>40</b>	52.70	54.60	62.20	66.15	69.35	29.50	31.20	48.30	59.40	69.50
<b>50</b>	54.00	54.10	57.90	64.80	68.90	30.65	35.05	58.90	70.00	74.65
<b>100</b>	63.15	62.25	69.60	73.15	76.80	37.90	50.40	80.10	90.55	94.10
<b>150</b>	69.85	73.65	76.25	83.15	86.45	48.30	62.05	91.05	97.10	98.15
<b>200</b>	79.70	81.05	86.30	89.55	92.15	56.20	71.60	95.30	99.15	99.60
<b>Moon and Weidner's QMLE with true number of factors (<math>m = 1</math>)</b>										
<b>40</b>	79.80	84.10	93.40	95.40	97.75	49.40	52.15	60.55	65.20	67.85
<b>50</b>	83.70	86.20	95.35	97.90	98.60	51.40	52.65	60.85	68.15	72.45
<b>100</b>	93.75	96.50	99.30	99.55	99.90	63.55	66.20	76.05	80.00	84.20
<b>150</b>	96.80	98.35	99.75	100.00	100.00	72.75	73.35	81.90	87.20	90.80
<b>200</b>	97.85	99.35	99.95	100.00	100.00	74.70	78.50	85.85	90.60	93.80

Notes: See notes to Table 1. CCEMG is based on (40) which features cross section averages of  $\mathbf{z}_{it} = (y_{it}, x_{it}, g_{it})'$ . QMLE estimator and MG estimator based on Song's individual estimates are computed from demeaned variables  $\hat{y}_{it}$  and  $\hat{x}_{it}$  defined in (41).

**Table 10.** Size and Power of estimating  $\phi$  in Experiment 16 (with regressors,  $\phi = 0.4$ ,  $m = 2$  and  $\rho_f = 0.6$ ).

(N,T)	Size (x100)					Power (x100)				
	40	50	100	150	200	40	50	100	150	200
<b>Fixed Effects estimates</b>										
40	97.30	99.45	99.95	100.00	100.00	84.35	90.30	97.70	99.60	99.80
50	98.50	99.30	100.00	100.00	100.00	85.85	92.40	99.00	99.85	99.90
100	99.50	99.85	100.00	100.00	100.00	92.05	95.50	99.75	100.00	100.00
150	99.85	99.95	100.00	100.00	100.00	93.95	96.05	99.75	100.00	100.00
200	99.95	99.95	100.00	100.00	100.00	94.15	97.40	99.95	100.00	100.00
<b>Dynamic CCEMG without bias correction</b>										
40	69.65	50.75	18.30	10.70	9.85	99.65	98.90	93.70	89.20	84.45
50	79.70	59.65	19.80	10.90	9.80	99.90	99.90	98.05	95.05	91.70
100	97.25	87.40	32.95	16.45	10.25	100.00	100.00	99.95	100.00	99.75
150	99.65	97.85	48.65	22.35	13.00	100.00	100.00	100.00	100.00	100.00
200	100.00	99.70	61.80	30.95	17.10	100.00	100.00	100.00	100.00	100.00
<b>Dynamic CCEMG with RMA bias correction</b>										
40	44.25	26.90	10.15	7.10	6.95	95.35	92.15	84.05	79.70	76.55
50	51.00	33.25	10.80	6.80	7.20	98.25	96.60	91.85	88.55	86.10
100	78.85	57.10	16.00	8.85	6.35	99.85	99.90	99.60	99.65	99.35
150	90.40	75.45	22.85	11.45	7.45	100.00	100.00	99.95	100.00	100.00
200	96.85	86.00	30.55	14.40	8.30	100.00	100.00	100.00	100.00	100.00
<b>Dynamic CCEMG with jackknife bias correction</b>										
40	14.30	12.35	11.20	8.60	9.15	19.55	27.65	48.30	56.45	62.45
50	15.85	11.50	10.75	8.95	8.00	21.80	32.15	56.20	65.90	72.20
100	22.15	17.95	11.90	9.25	6.45	33.70	49.60	85.30	92.65	96.35
150	29.05	21.40	12.10	10.30	8.00	39.85	64.95	96.00	98.60	99.65
200	34.00	26.40	14.25	10.40	7.30	46.35	72.00	98.85	99.80	100.00
<b>MG based on Song's individual estimates with 3 factors</b>										
40	53.25	34.75	9.40	8.40	6.85	95.95	95.20	86.30	80.10	78.55
50	59.95	40.95	10.05	7.75	7.40	99.00	98.25	92.10	89.50	86.20
100	88.60	70.25	17.50	9.40	6.90	99.90	99.95	99.75	99.60	99.05
150	96.60	83.85	23.60	11.15	8.05	100.00	100.00	100.00	100.00	100.00
200	99.35	94.20	32.80	14.00	8.85	100.00	100.00	100.00	100.00	100.00
<b>MG based on Song with true number of factors (<math>m = 2</math>)</b>										
40	44.00	28.40	8.30	7.65	6.50	95.90	94.30	85.15	78.80	77.25
50	50.15	34.35	9.10	8.20	7.45	98.65	97.95	91.50	89.20	85.15
100	79.15	60.35	15.20	8.55	6.65	99.95	100.00	99.65	99.30	99.15
150	92.00	76.05	20.75	10.30	7.55	100.00	100.00	100.00	99.90	100.00
200	97.45	88.45	28.15	12.35	8.25	100.00	100.00	100.00	100.00	100.00
<b>Moon and Weidner's QMLE with 3 factors</b>										
40	53.00	56.05	79.60	89.30	92.00	74.50	68.70	61.50	65.70	71.45
50	56.40	54.65	80.35	90.55	92.65	80.40	72.25	62.35	67.15	70.50
100	64.95	55.05	84.35	94.30	97.85	94.90	89.20	69.90	69.10	68.20
150	76.20	53.65	86.15	96.70	98.40	99.25	96.65	78.60	70.70	72.00
200	82.90	56.15	89.50	98.55	99.20	99.70	98.85	85.75	76.00	73.20
<b>Moon and Weidner's QMLE with true number of factors (<math>m = 2</math>)</b>										
40	51.55	59.30	84.60	90.25	94.30	64.75	59.10	61.95	67.70	70.95
50	52.10	57.55	83.95	92.05	93.50	69.95	63.20	62.40	64.70	70.05
100	45.70	55.20	89.10	96.20	98.45	87.25	78.35	63.25	63.10	66.15
150	45.35	56.55	94.40	98.65	99.65	93.35	88.25	64.35	64.40	67.20
200	46.50	59.80	97.60	99.75	99.85	97.05	92.85	67.70	67.40	70.10

Notes: See notes to Table 1. CCEMG is based on (40) which features cross section averages of  $\mathbf{z}_{it} = (y_{it}, x_{it}, g_{it})'$ . QMLE estimator and MG estimator based on Song's individual estimates are computed from demeaned variables  $\hat{y}_{it}$  and  $\hat{x}_{it}$  defined in (41).

**Table 11.** Size and Power of estimating  $\beta_0$  in Experiment 16 (with regressors,  $\phi = 0.4$ ,  $m = 2$  and  $\rho_f = 0.6$ ).

(N,T)	Size (x100)					Power (x100)				
	40	50	100	150	200	40	50	100	150	200
<b>Fixed Effects estimates</b>										
40	67.90	68.90	81.40	85.25	90.30	48.70	49.65	48.70	54.70	55.25
50	70.35	72.50	82.70	89.40	92.85	53.45	51.55	54.60	54.95	57.30
100	79.35	82.80	90.55	94.85	95.55	62.10	60.90	64.05	66.15	66.25
150	82.45	84.30	92.80	95.75	97.95	70.70	68.80	68.30	68.80	71.50
200	85.55	87.80	94.40	96.80	97.60	71.15	73.10	69.40	73.75	73.15
<b>Dynamic CCEMG without bias correction</b>										
40	6.00	5.85	6.50	6.65	6.25	35.55	46.10	73.80	87.35	90.75
50	5.60	5.90	5.75	6.80	6.30	43.25	54.30	83.70	92.25	95.70
100	5.40	6.85	5.45	5.50	5.70	69.00	81.85	98.65	99.85	99.95
150	7.20	5.50	5.70	6.60	4.95	85.00	95.30	99.95	100.00	100.00
200	7.30	6.35	6.30	5.20	4.95	93.25	98.05	100.00	100.00	100.00
<b>Dynamic CCEMG with RMA bias correction</b>										
40	5.80	6.15	6.50	6.65	6.50	29.25	40.05	68.95	84.30	89.55
50	5.75	6.30	7.25	6.50	6.80	35.50	47.60	80.00	90.15	94.85
100	5.40	5.70	5.55	6.05	6.40	58.50	76.15	97.65	99.65	99.85
150	7.45	5.55	6.90	6.80	5.45	74.05	89.85	99.90	100.00	99.95
200	6.60	6.00	6.50	5.95	5.70	86.75	96.45	100.00	100.00	100.00
<b>Dynamic CCEMG with jackknife bias correction</b>										
40	5.80	5.45	5.70	6.00	7.00	61.70	86.45	99.85	100.00	100.00
50	5.30	5.65	6.40	6.05	5.40	68.20	92.55	100.00	100.00	100.00
100	4.25	5.60	5.35	5.30	5.45	91.70	99.70	100.00	100.00	100.00
150	5.65	5.20	5.60	5.85	5.30	98.25	99.95	100.00	100.00	100.00
200	6.75	5.40	5.60	4.90	5.60	99.50	100.00	100.00	100.00	100.00
<b>MG based on Song's individual estimates with 3 factors</b>										
40	9.75	9.10	6.50	5.70	4.80	43.95	51.75	71.55	78.75	83.50
50	9.95	8.65	7.15	5.50	3.75	50.50	60.25	80.45	86.30	90.30
100	9.35	9.40	5.65	5.50	4.65	75.75	85.30	96.20	98.30	98.25
150	11.25	9.35	7.50	5.60	3.75	88.15	94.10	98.95	98.55	98.15
200	11.75	10.15	7.80	5.15	4.60	94.40	97.15	99.35	98.85	98.45
<b>MG based on Song with true number of factors (<math>m = 2</math>)</b>										
40	11.05	9.95	6.55	5.90	5.40	48.50	55.85	72.15	80.45	84.00
50	11.40	10.70	7.20	6.10	4.75	56.15	64.75	82.85	88.10	91.20
100	11.70	11.00	6.70	5.80	5.30	80.65	87.85	96.90	99.20	98.85
150	13.15	10.80	7.95	5.85	4.10	91.00	95.35	99.15	99.20	98.85
200	13.00	10.35	7.90	5.40	5.15	95.85	98.25	99.40	99.35	98.70
<b>Moon and Weidner's QMLE with 3 factors</b>										
40	39.00	41.70	51.75	57.60	61.90	37.10	41.60	58.60	70.20	75.20
50	40.95	44.35	53.40	62.00	67.10	41.20	45.70	64.80	70.50	77.15
100	56.90	58.40	71.95	76.85	80.30	46.80	56.70	76.40	86.80	90.15
150	67.05	70.40	80.15	86.05	89.30	54.20	66.55	87.60	93.10	95.80
200	77.05	78.50	86.30	91.25	93.00	59.70	73.85	93.25	97.80	98.40
<b>Moon and Weidner's QMLE with true number of factors (<math>m = 2</math>)</b>										
40	36.25	39.50	47.90	52.55	59.75	36.95	39.80	57.90	69.95	74.85
50	38.35	42.80	50.10	60.15	64.30	39.40	43.70	63.75	69.75	77.00
100	53.20	56.70	70.45	76.20	80.95	47.65	55.20	73.65	84.40	87.40
150	65.85	69.30	81.55	87.65	90.85	53.85	62.65	81.70	88.60	92.80
200	73.55	75.55	88.00	92.80	94.85	59.10	68.30	87.70	93.00	95.90

Notes: See notes to Table 1. CCEMG is based on (40) which features cross section averages of  $\mathbf{z}_{it} = (y_{it}, x_{it}, g_{it})'$ . QMLE estimator and MG estimator based on Song's individual estimates are computed from demeaned variables  $\hat{y}_{it}$  and  $\hat{x}_{it}$  defined in (41).



**Table 12.** Size and Power of estimating  $\phi$  in Experiment 18 (with regressors,  $\phi = 0.4$ ,  $m = 3$  and  $\rho_f = 0.6$ ).

(N,T)	Size (x100)					Power (x100)				
	40	50	100	150	200	40	50	100	150	200
<b>Fixed Effects estimates</b>										
<b>40</b>	99.40	99.85	100.00	100.00	100.00	91.55	95.75	99.70	99.95	99.90
<b>50</b>	99.60	100.00	100.00	100.00	100.00	93.80	97.35	99.85	100.00	100.00
<b>100</b>	100.00	100.00	100.00	100.00	100.00	97.25	98.40	100.00	100.00	100.00
<b>150</b>	99.90	100.00	100.00	100.00	100.00	98.05	99.25	100.00	100.00	100.00
<b>200</b>	100.00	100.00	100.00	100.00	100.00	97.75	99.35	100.00	100.00	100.00
<b>Dynamic CCEMG without bias correction</b>										
<b>40</b>	73.90	56.55	17.20	10.80	10.25	99.75	99.45	94.00	88.60	82.80
<b>50</b>	82.90	64.40	20.45	11.65	9.45	100.00	99.85	97.80	94.35	91.70
<b>100</b>	98.35	91.30	35.00	14.40	11.10	100.00	100.00	100.00	99.90	99.90
<b>150</b>	99.90	97.75	47.65	20.30	13.05	100.00	100.00	100.00	100.00	100.00
<b>200</b>	100.00	99.75	59.45	26.15	14.70	100.00	100.00	100.00	100.00	100.00
<b>Dynamic CCEMG with RMA bias correction</b>										
<b>40</b>	52.75	34.05	11.00	6.95	7.60	96.45	95.40	86.00	81.85	76.65
<b>50</b>	61.30	40.35	12.55	8.05	6.85	99.00	97.95	93.35	89.80	87.20
<b>100</b>	86.60	69.35	21.35	10.00	8.55	100.00	100.00	99.95	99.75	99.85
<b>150</b>	95.00	83.40	29.15	13.70	9.45	100.00	100.00	100.00	100.00	100.00
<b>200</b>	98.05	91.25	40.15	17.70	10.70	100.00	100.00	100.00	100.00	100.00
<b>Dynamic CCEMG with jackknife bias correction</b>										
<b>40</b>	15.45	13.15	11.15	9.05	9.35	18.70	27.20	45.70	55.00	60.05
<b>50</b>	17.65	14.55	10.30	9.45	8.30	20.05	31.05	56.05	66.20	71.30
<b>100</b>	21.45	17.30	12.00	10.50	6.70	32.40	49.35	84.20	91.75	96.55
<b>150</b>	28.00	21.90	11.40	8.85	6.85	41.35	61.00	95.90	98.80	99.50
<b>200</b>	34.60	25.30	13.85	10.20	7.80	45.50	71.90	98.65	99.70	99.95
<b>MG based on Song with true number of factors (<math>m = 3</math>)</b>										
<b>40</b>	44.45	24.95	7.50	6.40	7.90	94.50	90.75	73.55	65.95	62.80
<b>50</b>	50.00	32.45	7.90	7.80	8.50	97.50	95.85	83.20	77.70	73.90
<b>100</b>	82.40	58.70	9.55	6.45	8.65	99.95	99.85	98.95	97.80	96.60
<b>150</b>	94.10	77.90	15.10	8.55	6.60	100.00	100.00	99.90	99.85	99.85
<b>200</b>	98.65	90.05	24.25	9.50	7.20	100.00	100.00	100.00	100.00	100.00
<b>Moon and Weidner's QMLE with true number of factors (<math>m = 3</math>)</b>										
<b>40</b>	57.25	63.35	90.10	95.00	96.05	68.15	61.75	63.30	70.90	74.35
<b>50</b>	58.10	62.10	88.65	94.90	96.60	73.00	67.20	62.75	69.60	73.35
<b>100</b>	61.80	51.50	91.30	98.00	98.90	93.30	86.70	63.20	62.75	68.15
<b>150</b>	74.65	53.20	92.35	98.40	99.70	98.35	95.60	71.35	66.20	66.05
<b>200</b>	79.80	54.85	93.85	99.60	99.90	99.60	98.30	79.85	68.90	68.60

Notes: See notes to Table 1. CCEMG is based on (40) which features cross section averages of  $\mathbf{z}_{it} = (y_{it}, x_{it}, g_{it})'$ . QMLE estimator and MG estimator based on Song's individual estimates are computed from demeaned variables  $\dot{y}_{it}$  and  $\dot{x}_{it}$  defined in (41).

**Table 13.** Size and Power of estimating  $\beta_0$  in Experiment 18 (with regressors,  $\phi = 0.4$ ,  $m = 3$  and  $\rho_f = 0.6$ ).

(N,T)	Size (x100)					Power (x100)				
	40	50	100	150	200	40	50	100	150	200
	<b>Fixed Effects estimates</b>									
40	88.05	91.30	97.50	99.45	99.50	98.35	99.15	100.00	100.00	100.00
50	89.30	92.90	98.55	99.50	99.95	98.55	99.65	100.00	100.00	100.00
100	93.90	96.20	99.65	100.00	100.00	99.35	99.75	100.00	100.00	100.00
150	94.65	96.65	99.65	100.00	100.00	99.80	99.95	100.00	100.00	100.00
200	95.85	98.10	99.95	100.00	100.00	99.75	100.00	100.00	100.00	100.00
	<b>Dynamic CCEMG without bias correction</b>									
40	7.20	6.75	6.25	6.00	7.20	36.15	45.80	73.95	84.75	90.35
50	6.50	5.35	7.25	7.10	5.15	42.45	55.15	82.40	91.75	96.10
100	5.75	6.60	6.65	5.80	5.40	68.75	82.30	98.45	99.90	99.95
150	7.05	6.15	5.15	5.75	6.00	85.45	94.85	99.90	100.00	100.00
200	6.35	7.35	5.45	5.20	5.40	93.70	97.80	100.00	100.00	100.00
	<b>Dynamic CCEMG with RMA bias correction</b>									
40	7.15	6.80	6.10	6.50	6.85	28.90	38.55	69.70	81.05	89.05
50	6.05	7.15	6.40	7.05	5.15	35.55	48.80	77.60	90.25	95.65
100	6.50	6.50	6.70	6.05	5.60	58.10	76.55	97.75	99.70	99.80
150	6.55	6.15	5.55	5.75	5.90	76.20	90.00	99.70	100.00	100.00
200	6.75	7.15	5.35	5.50	5.30	86.30	95.65	99.95	100.00	100.00
	<b>Dynamic CCEMG with jackknife bias correction</b>									
40	5.90	6.20	5.55	6.50	6.90	59.95	85.85	99.85	100.00	100.00
50	5.10	6.85	5.70	6.85	5.50	66.95	90.30	100.00	100.00	100.00
100	5.75	5.40	5.45	5.30	5.75	91.30	99.30	100.00	100.00	100.00
150	5.05	5.30	5.00	5.35	5.85	97.50	99.95	100.00	100.00	100.00
200	5.55	5.90	4.95	5.30	5.45	99.45	100.00	100.00	100.00	100.00
	<b>MG based on Song with true number of factors (<math>m = 3</math>)</b>									
40	8.45	8.55	5.15	4.90	5.05	41.95	50.75	72.65	79.20	80.80
50	8.20	9.10	6.85	5.00	4.50	52.00	59.00	79.50	86.95	89.75
100	9.35	8.55	7.10	5.55	4.85	76.70	84.20	95.65	97.85	97.25
150	8.70	8.25	6.05	5.40	4.75	90.15	94.65	98.30	98.45	97.90
200	9.45	8.30	6.65	4.55	4.05	95.20	97.60	98.60	98.90	98.05
	<b>Moon and Weidner's QMLE with true number of factors (<math>m = 3</math>)</b>									
40	33.45	39.20	48.30	58.05	63.25	41.35	44.25	61.60	68.50	75.90
50	37.15	43.05	53.85	60.50	65.60	40.65	44.90	64.75	75.35	80.60
100	54.05	56.30	70.35	77.85	83.05	49.60	59.00	80.15	88.40	90.95
150	63.35	68.80	83.45	85.70	90.35	59.10	68.40	88.90	94.95	95.45
200	73.65	78.75	88.80	91.30	93.75	67.30	74.70	94.05	97.55	98.70

Notes: See notes to Table 1. CCEMG is based on (40) which features cross section averages of  $\mathbf{z}_{it} = (y_{it}, x_{it}, g_{it})'$ . QMLE estimator and MG estimator based on Song's individual estimates are computed from demeaned variables  $\dot{y}_{it}$  and  $\dot{x}_{it}$  defined in (41).

# A Mathematical Appendix

## A.1 Notations and Definitions

We begin by briefly summarizing the notations used in the paper, and introduce new notations which will prove useful in the proofs provided below. All vectors are represented by bold lower case letters and matrices are represented by bold upper case letters. We use  $\langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{a}'\mathbf{b}$  to denote the inner product (corresponding to the Euclidean norm) of vectors  $\mathbf{a}$  and  $\mathbf{b}$ .  $\|\mathbf{A}\|_1 \equiv \max_{1 \leq j \leq n} \sum_{i=1}^n |a_{ij}|$ , and  $\|\mathbf{A}\|_\infty \equiv \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}|$  denote the maximum absolute column and row sum norms of  $\mathbf{A} \in \mathbb{M}^{n \times n}$ , respectively, where  $\mathbb{M}^{n \times n}$  is the space of real-valued  $n \times n$  matrices.  $\|\mathbf{A}\| = \sqrt{\varrho(\mathbf{A}'\mathbf{A})}$  is the spectral norm of  $\mathbf{A}$ ,  $\varrho(\mathbf{A}) \equiv \max_{1 \leq i \leq n} \{|\lambda_i(\mathbf{A})|\}$  is the spectral radius of  $\mathbf{A}$ , and  $|\lambda_1(\mathbf{A})| \geq |\lambda_2(\mathbf{A})| \geq \dots \geq |\lambda_n(\mathbf{A})|$  are the eigenvalues of  $\mathbf{A}$ .  $Col(\mathbf{A})$  denotes the space spanned by the column vectors of  $\mathbf{A}$ . Note that  $\|\mathbf{a}\| = \sqrt{\varrho(\mathbf{a}'\mathbf{a})} = \sqrt{\mathbf{a}'\mathbf{a}}$  corresponds to the Euclidean length of vector  $\mathbf{a}$ .

Let

$$\mathbf{y}_i = \begin{pmatrix} y_{i,p_T+1} \\ y_{i,p_T+2} \\ \vdots \\ y_{iT} \end{pmatrix}_{T-p_T \times 1}, \quad \mathbf{y}_{i,-1} = \begin{pmatrix} y_{ip_T} \\ y_{i,p_T+1} \\ \vdots \\ y_{i,T-1} \end{pmatrix}_{T-p_T \times 1}, \quad \mathbf{X}_i = \begin{pmatrix} \mathbf{x}'_{i,p_T+1} \\ \mathbf{x}'_{i,p_T+2} \\ \vdots \\ \mathbf{x}'_{iT} \end{pmatrix}_{T-p_T \times k_x}, \quad \mathbf{X}_{i,-1} = \begin{pmatrix} \mathbf{x}'_{ip_T} \\ \mathbf{x}'_{i,p_T+1} \\ \vdots \\ \mathbf{x}'_{i,T-1} \end{pmatrix}_{T-p_T \times k_x},$$

$\boldsymbol{\tau}_{T-p_T} = (1, 1, \dots, 1)'$  is  $T - p_T \times 1$  vector of ones,  $\boldsymbol{\xi}_{it} = (y_{i,t-1}, \mathbf{x}'_{it}, \mathbf{x}'_{i,t-1})'$ ,

$$\boldsymbol{\Xi}_i = \begin{pmatrix} \boldsymbol{\xi}'_{i,p_T+1} \\ \boldsymbol{\xi}'_{i,p_T+2} \\ \vdots \\ \boldsymbol{\xi}'_{iT} \end{pmatrix}_{T-p_T \times 2k_x+1} = (\mathbf{y}_{i,-1}, \mathbf{X}_i, \mathbf{X}_{i,-1}), \quad \mathbf{F} = \begin{pmatrix} \mathbf{f}'_{p_T+1} \\ \mathbf{f}'_{p_T+2} \\ \vdots \\ \mathbf{f}'_T \end{pmatrix}_{T-p_T \times m}, \quad \text{and } \boldsymbol{\varepsilon}_i = \begin{pmatrix} \varepsilon_{i,p_T+1} \\ \varepsilon_{i,p_T+2} \\ \vdots \\ \varepsilon_{iT} \end{pmatrix}.$$

Using the above notations, model (1) can be written as

$$\mathbf{y}_i = c_{yi}\boldsymbol{\tau}_{T-p_T} + \phi_i\mathbf{y}_{i,-1} + \mathbf{X}_i\boldsymbol{\beta}_{0i} + \mathbf{X}_{i,-1}\boldsymbol{\beta}_{1i} + \mathbf{F}\boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_i,$$

or more compactly as

$$\mathbf{y}_i = c_{yi}\boldsymbol{\tau}_{T-p_T} + \boldsymbol{\Xi}_i\boldsymbol{\pi}_i + \mathbf{F}\boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_i, \quad (\text{A.1})$$

for  $i = 1, 2, \dots, N$ , where  $\boldsymbol{\pi}_i = (\phi_i, \boldsymbol{\beta}'_{0i}, \boldsymbol{\beta}'_{1i})'$ . Let also  $\mathbf{z}_{it} = (y_{it}, \boldsymbol{\omega}'_{it})'$ ,  $\bar{\mathbf{z}}_{wt} = (\bar{y}_{wt}, \bar{\boldsymbol{\omega}}'_{wt})' = \sum_{i=1}^N w_i \mathbf{z}_{it}$ ,

$$\bar{\mathbf{Q}}_w = \begin{pmatrix} 1 & \bar{\mathbf{z}}'_{w,p_T+1} & \bar{\mathbf{z}}'_{w,p_T} & \cdots & \bar{\mathbf{z}}'_{w,1} \\ 1 & \bar{\mathbf{z}}'_{w,p_T+2} & \bar{\mathbf{z}}'_{w,p_T+1} & \cdots & \bar{\mathbf{z}}'_{w,2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \bar{\mathbf{z}}'_{w,T} & \bar{\mathbf{z}}'_{w,T-1} & \cdots & \bar{\mathbf{z}}'_{w,T-p_T} \end{pmatrix}_{T-p_T \times (k+1)p_T+1}, \quad \text{and } \boldsymbol{\eta}_i = \begin{pmatrix} \sum_{\ell=p_T+1}^{\infty} \boldsymbol{\delta}'_{i\ell} \bar{\mathbf{z}}_{w,p_T+1-\ell} \\ \sum_{\ell=p_T+1}^{\infty} \boldsymbol{\delta}'_{i\ell} \bar{\mathbf{z}}_{w,p_T+2-\ell} \\ \vdots \\ \sum_{\ell=p_T+1}^{\infty} \boldsymbol{\delta}'_{i\ell} \bar{\mathbf{z}}_{w,T-\ell} \end{pmatrix}.$$

Model (A.1) can be equivalently written as (see also (22)),

$$\mathbf{y}_i = \boldsymbol{\Xi}_i\boldsymbol{\pi}_i + \bar{\mathbf{Q}}_w\mathbf{d}_i + \boldsymbol{\varepsilon}_i + \boldsymbol{\eta}_i + \boldsymbol{\vartheta}_i, \quad (\text{A.2})$$

where  $\mathbf{d}_i = (c_{yi}^*, \boldsymbol{\delta}'_{i0}, \boldsymbol{\delta}'_{i1}, \dots, \boldsymbol{\delta}'_{ip_T})'$ ,  $\boldsymbol{\delta}_i(L)$  is given by  $\boldsymbol{\delta}_i(L) = \mathbf{G}'(L)\boldsymbol{\gamma}_i = [\boldsymbol{\gamma}'_i(\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}'\boldsymbol{\Lambda}^{-1}(L)]'$ , see (23),

$$c_{yi}^* = c_{yi} - \delta_i'(1) \bar{\mathbf{c}}_{zw}, \text{ and}$$

$$\begin{aligned} \boldsymbol{\vartheta}_i &= c_{yi} \boldsymbol{\tau} + \mathbf{F} \boldsymbol{\gamma}_i - \bar{\mathbf{Q}}_w \mathbf{d}_i - \boldsymbol{\eta}_i \\ &= \mathbf{F} \boldsymbol{\gamma}_i - \tilde{\mathbf{Z}}_w \boldsymbol{\delta}_i(L), \end{aligned} \quad (\text{A.3})$$

in which

$$\tilde{\mathbf{Z}}_w = \bar{\mathbf{Z}}_w - \boldsymbol{\tau}_{T-p_T} \mathbf{c}'_{zw}, \quad \bar{\mathbf{Z}}_w = \begin{pmatrix} \bar{\mathbf{z}}'_{w,p_T+1} \\ \bar{\mathbf{z}}'_{w,p_T+2} \\ \vdots \\ \bar{\mathbf{z}}'_{w,T} \end{pmatrix}, \text{ and } \mathbf{c}_{zw} = \sum_{i=1}^N w_i (\mathbf{I}_{k+1} - \mathbf{A}_i)^{-1} \mathbf{c}_{zi}.$$

Note that the individual elements of  $\boldsymbol{\vartheta}_i = (\vartheta_{i,p_T+1}, \vartheta_{i,p_T+2}, \dots, \vartheta_{i,T})'$  are  $O_p(N^{-1/2})$  uniformly across all  $i$  and  $t$ .

Define also the following projection matrices

$$\mathbf{P}_h = \mathbf{H}_w (\mathbf{H}_w' \mathbf{H}_w)^+ \mathbf{H}_w', \text{ and } \mathbf{M}_h = \mathbf{I}_{T-p_T} - \mathbf{H}_w (\mathbf{H}_w' \mathbf{H}_w)^+ \mathbf{H}_w', \quad (\text{A.4})$$

in which

$$\mathbf{H}_w = \begin{pmatrix} 1 & \mathbf{h}'_{w,p_T+1} & \mathbf{h}'_{wp_T} & \cdots & \mathbf{h}'_{w1} \\ 1 & \mathbf{h}'_{w,p_T+2} & \mathbf{h}'_{w,p_T+1} & \cdots & \mathbf{h}'_{w2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \mathbf{h}'_{w,T} & \mathbf{h}'_{w,T-1} & \cdots & \mathbf{h}'_{w,T-p_T} \end{pmatrix},$$

and  $\mathbf{h}_{wt} = \boldsymbol{\Psi}_w(L) \mathbf{f}_t + \mathbf{c}_{zw}$ , where

$$\boldsymbol{\Psi}_w(L) = \sum_{i=1}^N w_i (\mathbf{I}_{k+1} - \mathbf{A}_i L)^{-1} \mathbf{A}_{0,i}^{-1} \mathbf{C}_i.$$

Furthermore, let  $\bar{\mathbf{V}}_w = \bar{\mathbf{Q}}_w - \mathbf{H}_w$ , and note that

$$\bar{\mathbf{V}}_w = \begin{pmatrix} 0 & \bar{\boldsymbol{\nu}}'_{w,p_T+1} & \bar{\boldsymbol{\nu}}'_{wp_T} & \cdots & \bar{\boldsymbol{\nu}}'_{w1} \\ 0 & \bar{\boldsymbol{\nu}}'_{w,p_T+2} & \bar{\boldsymbol{\nu}}'_{w,p_T+1} & \cdots & \bar{\boldsymbol{\nu}}'_{w2} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & \bar{\boldsymbol{\nu}}'_{wT} & \bar{\boldsymbol{\nu}}'_{w,T-1} & \cdots & \bar{\boldsymbol{\nu}}'_{w,T-p_T} \end{pmatrix}, \quad \bar{\boldsymbol{\nu}}_{wt} = \sum_{i=1}^N w_i (\mathbf{I}_{k+1} - \mathbf{A}_i L)^{-1} \mathbf{A}_{0,i}^{-1} \mathbf{e}_{it},$$

and  $\mathbf{H}_w = \tilde{\mathbf{F}} \tilde{\boldsymbol{\Lambda}}_w$ , where

$$\begin{aligned} \tilde{\mathbf{F}} &= \begin{pmatrix} 1 & \mathbf{f}'_{p_T+1} & \mathbf{f}'_{p_T} & \cdots & \mathbf{f}'_1 \\ 1 & \mathbf{f}'_{p_T+2} & \mathbf{f}'_{p_T+1} & \cdots & \mathbf{f}'_2 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \mathbf{f}'_T & \mathbf{f}'_{T-1} & \cdots & \mathbf{f}'_{T-p_T} \end{pmatrix}, \\ \tilde{\boldsymbol{\Lambda}}_w &= \begin{pmatrix} 1 & \mathbf{c}'_{zw} & \mathbf{c}'_{zw} & \cdots & \mathbf{c}'_{zw} \\ \mathbf{0}_{m \times 1} & \boldsymbol{\Lambda}'_w(L) & \mathbf{0}_{m \times k+1} & \cdots & \mathbf{0}_{m \times k+1} \\ \mathbf{0}_{m \times 1} & \mathbf{0}_{m \times k+1} & \boldsymbol{\Lambda}'_w(L) & & \mathbf{0}_{m \times k+1} \\ \vdots & \vdots & & \ddots & \vdots \\ \mathbf{0}_{m \times 1} & \mathbf{0}_{m \times k+1} & \mathbf{0}_{m \times k+1} & & \boldsymbol{\Lambda}'_w(L) \end{pmatrix}, \text{ and } \boldsymbol{\Lambda}_w(L) = \sum_{i=1}^N w_i (\mathbf{I}_{k+1} - \mathbf{A}_i L)^{-1} \mathbf{A}_{0,i}^{-1} \mathbf{C}_i. \end{aligned}$$

We also define

$$\mathbf{S}_{(1+2k_x) \times (1+2k_x)} = \begin{pmatrix} 1 & \mathbf{0}_{1 \times k_x} & \mathbf{0}_{1 \times k_x} \\ \mathbf{0}_{k_x \times 1} & \mathbf{0}_{k_x \times k_x} & \mathbf{I}_{k_x} \\ \mathbf{0}_{k_x \times 1} & \mathbf{I}_{k_x} & \mathbf{0}_{k_x \times k_x} \end{pmatrix}, \quad (\text{A.5})$$

$\boldsymbol{\xi}_{it}^* = (y_{i,t-1}, \mathbf{x}'_{i,t-1}, \mathbf{x}'_{it})'$ , and note that  $\boldsymbol{\xi}_{it} = \mathbf{S}' \boldsymbol{\xi}_{it}^*$ , and  $\boldsymbol{\Xi}_i = \boldsymbol{\Xi}_i^* \mathbf{S}$ , where  $\boldsymbol{\Xi}_i^* = (\boldsymbol{\xi}_{i,p_T+1}^*, \boldsymbol{\xi}_{i,p_T+2}^*, \dots, \boldsymbol{\xi}_{iT}^*)'$ . Individual elements of  $\boldsymbol{\xi}_{it}$  are also denoted as  $\xi_{ist}$  for  $s = 1, 2, \dots, 2k_x + 1$ , and the vector of observations on  $\boldsymbol{\xi}_{ist}$  is

$$\boldsymbol{\xi}_{is\cdot} = \begin{pmatrix} \xi_{i,s,p_T+1} \\ \vdots \\ \xi_{isT} \end{pmatrix}.$$

Recall that the panel data model (1)-(3) can be written as the VAR model (6) in  $\mathbf{z}_{it} = (y_{it}, \mathbf{x}'_{it}, \mathbf{g}'_{it})'$ . Hence we have

$$\mathbf{z}_{it} = \sum_{\ell=0}^{\infty} \mathbf{A}_i^{\ell} (\mathbf{c}_{zi} + \mathbf{A}_{0i}^{-1} \mathbf{C}_i \mathbf{f}_{t-\ell} + \mathbf{A}_{0i}^{-1} \mathbf{e}_{i,t-\ell}),$$

and

$$\boldsymbol{\xi}_{it}^* = \begin{pmatrix} y_{i,t-1} \\ \mathbf{x}_{i,t-1} \\ \mathbf{x}_{it} \end{pmatrix} = \begin{pmatrix} \mathbf{S}'_{yx} \mathbf{z}_{i,t-1} \\ \mathbf{S}'_x \mathbf{z}_{it} \end{pmatrix} = \mathbf{c}_{\xi^*i} + \boldsymbol{\Psi}_{\xi_i}(L) (\mathbf{C}_i \mathbf{f}_t + \mathbf{e}_{it}),$$

where

$$\mathbf{S}'_{yx} = \begin{pmatrix} 1 & \mathbf{0}_{1 \times k_x} & \mathbf{0}_{1 \times k_g} \\ \mathbf{0}_{k_x \times 1} & \mathbf{I}_{k_x} & \mathbf{0}_{k_x \times k_g} \end{pmatrix}, \quad \mathbf{S}'_x = \begin{pmatrix} \mathbf{0}_{k_x \times 1} & \mathbf{I}_{k_x} & \mathbf{0}_{k_x \times k_g} \end{pmatrix},$$

$\mathbf{c}_{\xi^*i} = \boldsymbol{\Psi}_{\xi_i}(L) (\mathbf{S}_{yx}, \mathbf{S}_x)' \mathbf{c}_{zi}$ , and

$$\boldsymbol{\Psi}_{\xi_i}(L) = \begin{pmatrix} \mathbf{0}_{k_x+1 \times k+1} \\ \mathbf{S}'_x \end{pmatrix} \mathbf{A}_{0i}^{-1} + \begin{pmatrix} \mathbf{S}'_{yx} (\mathbf{I}_{k+1} - \mathbf{A}_i L)^{-1} L \\ \mathbf{S}'_x [(\mathbf{I}_{k+1} - \mathbf{A}_i L)^{-1} - \mathbf{I}_{k+1}] \end{pmatrix} \mathbf{A}_{0i}^{-1}. \quad (\text{A.6})$$

## A.2 Statement of Lemmas

**Lemma A.1** Let  $\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_{s_N})$  and  $\mathbf{B} = (\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_{s_N})$  be  $r_N \times s_N$  random matrices, and  $r_N$  and  $s_N$  are deterministic sequences nondecreasing in  $N$ . Suppose also that  $\|\mathbf{a}_{\ell}\| = O_p(r_N^{1/2})$  and  $\|\mathbf{b}_{\ell}\| = O_p(r_N^{1/2} N^{-1/2})$ , uniformly in  $\ell$ , for  $\ell = 1, 2, \dots, s_N$ . Then for any  $\boldsymbol{\alpha}_{A,1}, \boldsymbol{\alpha}_{A,2} \in \text{Col}(\mathbf{A})$  for which there exist vectors  $\mathbf{c}_1$  and  $\mathbf{c}_2$  such that  $\boldsymbol{\alpha}_{A,1} = \mathbf{A} \mathbf{c}_1$ ,  $\boldsymbol{\alpha}_{A,2} = \mathbf{A} \mathbf{c}_2$ ,  $\|\mathbf{c}_1\|_{\infty} < K$  and  $\|\mathbf{c}_2\|_{\infty} < K$ , where the constant  $K < \infty$  does not depend on  $N$ , we have

$$\|\mathbf{M}_{A+B} \boldsymbol{\alpha}_{A,1}\| = O_p\left(\frac{s_N \sqrt{r_N}}{\sqrt{N}}\right), \quad (\text{A.7})$$

and

$$\langle \mathbf{M}_{A+B} \boldsymbol{\alpha}_{A,1}, \mathbf{M}_{A+B} \boldsymbol{\alpha}_{A,2} \rangle = \boldsymbol{\alpha}'_{A,1} \mathbf{M}_{A+B} \boldsymbol{\alpha}_{A,2} = O_p\left(\frac{s_N^2 r_N}{N}\right) \quad (\text{A.8})$$

where  $\mathbf{M}_{A+B}$  is orthogonal projection matrix that projects onto the orthogonal complement of  $\text{Col}(\mathbf{A} + \mathbf{B})$ .

**Lemma A.2** Suppose Assumptions 1-5 and 7 hold and  $(N, T, p_T) \xrightarrow{j} \infty$ . Then

$$\frac{1}{T} \sum_{t=1}^T y_{i,t-1} \varepsilon_{it} \xrightarrow{p} 0, \text{ uniformly in } i \quad (\text{A.9})$$

$$\frac{1}{T} \sum_{t=1}^T \boldsymbol{\omega}_{i,t-s} \varepsilon_{it} \xrightarrow{p} \mathbf{0}_{k \times 1}, \text{ uniformly in } i, \quad (\text{A.10})$$

and, if also  $p_T^3/T \rightarrow \kappa$  for some constant  $0 < \kappa < \infty$ ,

$$\frac{1}{T} \sum_{t=1}^T \mathbf{h}_{w,t-q} \varepsilon_{it} = O_p \left( T^{-1/2} \right), \text{ uniformly in } i \text{ and } q, \quad (\text{A.11})$$

for  $i = 1, 2, \dots, N$ ,  $q = 1, 2, \dots, p_T$ , and  $s = 0, 1$ . The same results hold when  $\varepsilon_{it}$  is replaced by  $\eta_{it}$  and  $\vartheta_{it}$ .

**Lemma A.3** Suppose Assumptions 1-5 and 7 hold and  $(N, T, p_T) \xrightarrow{j} \infty$  such that  $p_T^3/T \rightarrow \kappa$ ,  $0 < \kappa < \infty$ . Then

$$\frac{\boldsymbol{\Xi}'_i \mathbf{M}_h \boldsymbol{\Xi}_i}{T} \xrightarrow{p} \boldsymbol{\Sigma}_{i\xi} \text{ uniformly in } i, \quad (\text{A.12})$$

and

$$\frac{\boldsymbol{\Xi}'_i \mathbf{M}_h \mathbf{F}}{T} \xrightarrow{p} \mathbf{Q}_{if} \text{ uniformly in } i, \quad (\text{A.13})$$

where  $\boldsymbol{\Sigma}_{i\xi}$  is positive definite and given by

$$\boldsymbol{\Sigma}_{i\xi} = \boldsymbol{\Omega}_{\Psi\xi i} + \boldsymbol{\Omega}_{fi}, \quad (\text{A.14})$$

and

$$\mathbf{Q}_{if} = \text{cov} [\mathbf{S}' \boldsymbol{\Psi}_{\xi i}(L) \mathbf{C}_i^* \mathbf{f}_t, \mathbf{C}_i^* \mathbf{f}_t], \quad (\text{A.15})$$

in which

$$\boldsymbol{\Omega}_{\Psi\xi i} = \text{Var} [\mathbf{S}' \boldsymbol{\Psi}_{\xi i}(L) \mathbf{e}_{it}], \quad \boldsymbol{\Omega}_{fi} = \text{Var} [\mathbf{S}' \boldsymbol{\Psi}_{\xi i}(L) \mathbf{C}_i^* \mathbf{f}_t], \quad (\text{A.16})$$

$\mathbf{C}_i^* = \mathbf{M}_c \mathbf{C}_i$ ,  $\mathbf{M}_c = \mathbf{I}_{k+1} - \mathbf{C} \mathbf{C}^+$  is orthogonal projector onto the orthogonal complement of  $\text{Col}(\mathbf{C})$ ,  $\boldsymbol{\Psi}_{\xi i}(L) = \sum_{\ell=0}^{\infty} \boldsymbol{\Psi}_{\xi i \ell} L^\ell$  is defined in (A.6), selection matrix  $\mathbf{S}$  is defined in (A.5) and  $\mathbf{e}_{it} = (\varepsilon_{it}, \mathbf{v}'_{it})'$ . When factors are serially uncorrelated, then  $\boldsymbol{\Omega}_{fi} = \sum_{\ell=0}^{\infty} \mathbf{S}' \boldsymbol{\Psi}_{\xi i \ell} (\mathbf{C}_i^* \boldsymbol{\Omega}_f \mathbf{C}_i^{*'}) \boldsymbol{\Psi}'_{\xi i \ell} \mathbf{S}$  and  $\mathbf{Q}_{if} = \mathbf{S}' \boldsymbol{\Psi}_{\xi i 0} (\mathbf{C}_i^* \boldsymbol{\Omega}_f \mathbf{C}_i^{*'})$ , where  $\boldsymbol{\Omega}_f = \text{Var}(\mathbf{f}_t)$ .

**Lemma A.4** Suppose Assumptions 1-5 and 7 hold and  $(N, T, p_T) \xrightarrow{j} \infty$  such that  $p_T^3/T \rightarrow \kappa$  for some constant  $0 < \kappa < \infty$ . Then,

$$\frac{\boldsymbol{\Xi}'_i \mathbf{M}_h \boldsymbol{\varepsilon}_i}{T} \xrightarrow{p} \mathbf{0}_{2k_x+1 \times 1}, \text{ uniformly in } i, \quad (\text{A.17})$$

$$\frac{\boldsymbol{\Xi}'_i \mathbf{M}_h \boldsymbol{\eta}_i}{T} \xrightarrow{p} \mathbf{0}_{2k_x+1 \times 1}, \text{ uniformly in } i, \quad (\text{A.18})$$

and

$$\frac{\boldsymbol{\Xi}'_i \mathbf{M}_h \boldsymbol{\vartheta}_i}{T} \xrightarrow{p} \mathbf{0}_{2k_x+1 \times 1}, \text{ uniformly in } i. \quad (\text{A.19})$$

**Lemma A.5** Suppose Assumptions 1-5 hold and unobserved common factors are serially uncorrelated. Then, as  $(N, T, p_T) \xrightarrow{j} \infty$ , we have

$$\frac{1}{N} \sum_{i=1}^N \boldsymbol{\Sigma}_{i\xi}^{-1} \frac{\boldsymbol{\Xi}'_i \mathbf{M}_h \mathbf{F}}{T} \boldsymbol{\eta}_{\gamma i} \xrightarrow{p} \mathbf{0}_{2k_x+1 \times 1}. \quad (\text{A.20})$$

**Lemma A.6** Suppose Assumptions 1-5 hold and  $(N, T, p_T) \xrightarrow{j} \infty$  such that  $p_T^2/T \rightarrow 0$ . Then,

$$\sqrt{N} \frac{\Xi'_i \bar{\mathbf{M}}_q \Xi_i}{T} - \sqrt{N} \frac{\Xi'_i \mathbf{M}_h \Xi_i}{T} \xrightarrow{p} \mathbf{0}_{2k_x+1 \times 2k_x+1} \text{ uniformly in } i, \quad (\text{A.21})$$

$$\sqrt{N} \frac{\Xi'_i \bar{\mathbf{M}}_q \varepsilon_i}{T} - \sqrt{N} \frac{\Xi'_i \mathbf{M}_h \varepsilon_i}{T} \xrightarrow{p} \mathbf{0}_{2k_x+1 \times 1} \text{ uniformly in } i, \quad (\text{A.22})$$

$$\sqrt{N} \frac{\Xi'_i \bar{\mathbf{M}}_q \mathbf{F}}{T} - \sqrt{N} \frac{\Xi'_i \mathbf{M}_h \mathbf{F}}{T} \xrightarrow{p} \mathbf{0}_{2k_x+1 \times m} \text{ uniformly in } i. \quad (\text{A.23})$$

$$\frac{\Xi'_i \bar{\mathbf{M}}_q \boldsymbol{\eta}_i}{T} - \frac{\Xi'_i \mathbf{M}_h \boldsymbol{\eta}_i}{T} \xrightarrow{p} \mathbf{0}_{2k_x+1 \times 1}, \text{ uniformly in } i, \quad (\text{A.24})$$

and

$$\frac{\Xi'_i \bar{\mathbf{M}}_q \boldsymbol{\vartheta}_i}{T} - \frac{\Xi'_i \mathbf{M}_h \boldsymbol{\vartheta}_i}{T} \xrightarrow{p} \mathbf{0}_{2k_x+1 \times 1}, \text{ uniformly in } i. \quad (\text{A.25})$$

**Lemma A.7** Suppose Assumptions 1-5 hold and  $(N, T, p_T) \xrightarrow{j} \infty$  such that  $N/T \rightarrow \varkappa$ , for some  $0 < \varkappa < \infty$ , and  $p_T^2/T \rightarrow 0$ . Then,

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{\Xi'_i \mathbf{M}_h \varepsilon_i}{T} \xrightarrow{p} \mathbf{0}_{2k_x+1 \times 1}. \quad (\text{A.26})$$

### A.3 Proofs of Lemmas

**Proof of Lemma A.1.** Hilbert projection theorem (see Rudin, 1987) implies

$$\|\mathbf{M}_{A+B} \boldsymbol{\alpha}_{A,1}\| \leq \|\boldsymbol{\alpha}_{A,1} - \boldsymbol{\beta}_{A+B}\|, \quad (\text{A.27})$$

for any vector  $\boldsymbol{\beta}_{A+B} \in \text{Col}(\mathbf{A} + \mathbf{B})$ . Consider the following choice of  $\boldsymbol{\beta}_{A+B}$ ,

$$\boldsymbol{\beta}_{A+B} = \sum_{\ell=1}^{s_N} \mathbf{P}_{a_\ell+b_\ell} \mathbf{a}_\ell c_{1\ell}, \quad (\text{A.28})$$

where  $\mathbf{P}_{a_\ell+b_\ell}$  is orthogonal projector onto  $\text{Col}(\mathbf{a}_\ell + \mathbf{b}_\ell)$ , and  $c_{1\ell}$ , for  $\ell = 1, 2, \dots, s_N$  are elements of vector  $\mathbf{c}_1$ . Using  $\boldsymbol{\alpha}_{A,1} = \mathbf{A} \mathbf{c}_1 = \sum_{\ell=1}^{s_N} a_\ell c_{1\ell}$ , (A.27) with  $\boldsymbol{\beta}_{A+B}$  given by (A.28) can be written as

$$\|\mathbf{M}_{A+B} \boldsymbol{\alpha}_{A,1}\| \leq \left\| \sum_{\ell=1}^{s_N} \mathbf{a}_\ell c_{1\ell} - \sum_{\ell=1}^{s_N} \mathbf{P}_{a_\ell+b_\ell} \mathbf{a}_\ell c_{1\ell} \right\|.$$

Using now the triangle inequality, we obtain

$$\begin{aligned} \|\mathbf{M}_{A+B} \boldsymbol{\alpha}_{A,1}\| &\leq \sum_{\ell=1}^{s_N} \|\mathbf{a}_\ell c_{1\ell} - \mathbf{P}_{a_\ell+b_\ell} \mathbf{a}_\ell c_{1\ell}\| \\ &\leq \sum_{\ell=1}^{s_N} |c_{1\ell}| \|\mathbf{a}_\ell - \mathbf{P}_{a_\ell+b_\ell} \mathbf{a}_\ell\| \end{aligned} \quad (\text{A.29})$$

Next, we establish an upper bound to  $\|\mathbf{a}_\ell - \mathbf{P}_{a_\ell+b_\ell} \mathbf{a}_\ell\|$ . Consider the triangle given by  $\mathbf{a}_\ell$ ,  $\mathbf{P}_{a_\ell+b_\ell} \mathbf{a}_\ell$  and  $\mathbf{a}_\ell + \mathbf{b}_\ell$ . Hilbert projection theorem (see Rudin, 1987) implies

$$\|\mathbf{a}_\ell - \mathbf{P}_{a_\ell+b_\ell} \mathbf{a}_\ell\| \leq \|\mathbf{a}_\ell - (\mathbf{a}_\ell + \mathbf{b}_\ell) \gamma\|,$$

for any scalar  $\gamma$  and setting  $\gamma = 1$  we have

$$\begin{aligned}\|\mathbf{a}_\ell - \mathbf{P}_{\mathbf{a}_\ell + \mathbf{b}_\ell} \mathbf{a}_\ell\| &\leq \|\mathbf{a}_\ell - \mathbf{a}_\ell + \mathbf{b}_\ell\|, \\ &\leq \|\mathbf{b}_\ell\|, \\ &= O_p\left(r_N^{1/2} N^{-1/2}\right).\end{aligned}$$

Using this result in (A.29) and noting that  $|c_{1\ell}| < K$  by assumption, it follows that

$$\|\mathbf{M}_{A+B} \boldsymbol{\alpha}_{A,1}\| = O_p\left(\frac{s_N r_N^{1/2}}{N^{1/2}}\right),$$

as desired.

Consider now the inner product of vectors  $\mathbf{M}_{A+B} \boldsymbol{\alpha}_{A,1}$  and  $\mathbf{M}_{A+B} \boldsymbol{\alpha}_{A,2}$ . Using Cauchy-Schwarz inequality, we obtain

$$|\boldsymbol{\alpha}'_{A,1} \mathbf{M}_{A+B} \boldsymbol{\alpha}_{A,2}| = |(\mathbf{M}_{A+B} \boldsymbol{\alpha}_{A,1})' (\mathbf{M}_{A+B} \boldsymbol{\alpha}_{A,2})| \leq \|\mathbf{M}_{A+B} \boldsymbol{\alpha}_{A,1}\| \|\mathbf{M}_{A+B} \boldsymbol{\alpha}_{A,2}\|.$$

But (A.7) implies that both  $\|\mathbf{M}_{A+B} \boldsymbol{\alpha}_{A,1}\|$  and  $\|\mathbf{M}_{A+B} \boldsymbol{\alpha}_{A,2}\|$  are  $O_p\left(s_N \sqrt{r_N}/\sqrt{N}\right)$ . These results establish (A.8), as desired. ■

**Proof of Lemma A.2.** Note that all processes,  $\varepsilon_{it}$ ,  $\eta_{it}$ ,  $\vartheta_{it}$ ,  $y_{it}$ ,  $\boldsymbol{\omega}_{it}$  and  $\mathbf{h}_{wt}$ , are stationary with absolutely summable autocovariances and their cross products are ergodic in mean. Lemma A.2 can be established in the same way as Lemma 1 in Chudik and Pesaran (2011) by applying a mixingale weak law. ■

**Proof of Lemma A.3.** Lemma (A.3) can be established in a similar way as Lemma A.5 in Chudik, Pesaran, and Tosetti (2011) and by observing that  $\mathbf{M}_h$  is asymptotically the orthogonal complement of the space spanned by  $\mathbf{Cf}_t$ . ■

**Proof of Lemma A.4.** Let us denote the individual columns of  $\boldsymbol{\Xi}_i$  as  $\boldsymbol{\xi}_{is}$ , for  $s = 1, 2, \dots, 2k+1$ , and define the scaled vectors  $\boldsymbol{\xi}_{is}^\diamond = T^{-1/2} \boldsymbol{\xi}_{is}$ , and  $\boldsymbol{\varepsilon}_i^\diamond = T^{-1/2} \boldsymbol{\varepsilon}_i$ . Since the individual elements of  $\boldsymbol{\xi}_{is}$  and  $\boldsymbol{\varepsilon}_i$  are uniformly  $O_p(1)$ , we have  $\|\boldsymbol{\xi}_{is}\| = O_p(T^{1/2})$ ,  $\|\boldsymbol{\varepsilon}_i\| = O_p(T^{1/2})$  and therefore  $\|\boldsymbol{\xi}_{is}^\diamond\| = O_p(1)$  and  $\|\boldsymbol{\varepsilon}_i^\diamond\| = O_p(1)$ . Now consider the inner product

$$\langle \mathbf{M}_h \boldsymbol{\xi}_{is}^\diamond, \mathbf{M}_h \boldsymbol{\varepsilon}_i^\diamond \rangle = \langle \boldsymbol{\xi}_{is}^\diamond, \boldsymbol{\varepsilon}_i^\diamond \rangle + \langle \mathbf{P}_h \boldsymbol{\xi}_{is}^\diamond, \mathbf{P}_h \boldsymbol{\varepsilon}_i^\diamond \rangle, \quad (\text{A.30})$$

where  $\langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{a}' \mathbf{b}$  denotes the inner product of vectors  $\mathbf{a}$  and  $\mathbf{b}$ , and  $\mathbf{P}_h = \mathbf{H}_w (\mathbf{H}_w' \mathbf{H}_w)^+ \mathbf{H}_w'$  is the orthogonal projection matrix that projects onto the column space of  $\mathbf{H}_w$ . Consider the probability limits of the elements in (A.30) as  $(N, T, p_T) \xrightarrow{j} \infty$  such that  $p_T^3/T \rightarrow \varkappa$  for some constant  $0 < \varkappa < \infty$ . (A.9) and (A.10) of Lemma A.2 establish that

$$\langle \boldsymbol{\xi}_{is}^\diamond, \boldsymbol{\varepsilon}_i^\diamond \rangle \xrightarrow{p} 0, \text{ for } s = 1, 2, \dots, 2k+1. \quad (\text{A.31})$$

Consider the Euclidean norm of the second term of (A.30). Using Cauchy-Schwarz inequality we obtain the following upper bound,

$$\|\langle \mathbf{P}_h \boldsymbol{\xi}_{is}^\diamond, \mathbf{P}_h \boldsymbol{\varepsilon}_i^\diamond \rangle\| \leq \|\mathbf{P}_h \boldsymbol{\xi}_{is}^\diamond\| \|\mathbf{P}_h \boldsymbol{\varepsilon}_i^\diamond\|, \quad (\text{A.32})$$

where (by Pythagoras' theorem)<sup>10</sup>

$$\|\mathbf{P}_h \boldsymbol{\xi}_{is}^\diamond\| \leq \|\boldsymbol{\xi}_{is}^\diamond\| = O_p(1). \quad (\text{A.33})$$

<sup>10</sup>Let  $\mathbf{M}_h = (\mathbf{I}_{T-p_T} - \mathbf{P}_h)$  and note that  $\boldsymbol{\xi}_{is}^\diamond = \mathbf{M}_h \boldsymbol{\xi}_{is}^\diamond + \mathbf{P}_h \boldsymbol{\xi}_{is}^\diamond$ . Vectors  $\mathbf{M}_h \boldsymbol{\xi}_{is}^\diamond$  and  $\mathbf{P}_h \boldsymbol{\xi}_{is}^\diamond$  are orthogonal and therefore  $\|\mathbf{M}_h \boldsymbol{\xi}_{is}^\diamond + \mathbf{P}_h \boldsymbol{\xi}_{is}^\diamond\|^2 = \|\mathbf{M}_h \boldsymbol{\xi}_{is}^\diamond\|^2 + \|\mathbf{P}_h \boldsymbol{\xi}_{is}^\diamond\|^2$ . It now follows that  $\|\boldsymbol{\xi}_{is}^\diamond\|^2 = \|\mathbf{M}_h \boldsymbol{\xi}_{is}^\diamond\|^2 + \|\mathbf{P}_h \boldsymbol{\xi}_{is}^\diamond\|^2$ , but since  $\|\mathbf{M}_h \boldsymbol{\xi}_{is}^\diamond\|^2 \geq 0$ , we obtain  $\|\boldsymbol{\xi}_{is}^\diamond\|^2 \geq \|\mathbf{P}_h \boldsymbol{\xi}_{is}^\diamond\|^2$ .



Now we will establish convergence of  $\|\mathbf{P}_h \boldsymbol{\varepsilon}_i^\diamond\|$  in probability. By spectral theorem there exists a unitary matrix  $\mathbf{V}$  such that

$$\mathbf{V}' \frac{\mathbf{H}_w' \mathbf{H}_w}{T} \mathbf{V} = \begin{pmatrix} \mathbf{D} & \mathbf{0}_{r_c p_T + 1 \times (k+1-r_c)p_T} \\ \mathbf{0}_{(k+1-r_c)p_T \times r_c p_T + 1} & \mathbf{0}_{(k+1-r_c)p_T \times (k+1-r_c)p_T} \end{pmatrix}, \quad (\text{A.34})$$

where  $\mathbf{D}$  is  $r_c p_T + 1$  dimensional diagonal matrix with strictly positive diagonal elements and  $r_c = \text{rank}(\mathbf{C})$ . Also by assumption  $\mathbf{f}_t$  is a stationary process with absolute summable autocovariances, and so is  $\mathbf{h}_{wt}$ . Furthermore,  $\mathbf{H}_w' \mathbf{H}_w / T = O_p(1)$  as well as the diagonal elements of  $\mathbf{D}$  have nonzero (and finite) probability limits. Partition unitary matrix  $\mathbf{V} = (\mathbf{V}_1, \mathbf{V}_2)$  so that  $T^{-1} \mathbf{V}_1' \mathbf{H}_w' \mathbf{H}_w \mathbf{V}_1 = \mathbf{D}$  and define  $\mathbf{U}_1 = T^{-1/2} \mathbf{H}_w \mathbf{V}_1 \mathbf{D}^{-1/2}$ . Note that  $\mathbf{U}_1$  is orthonormal basis of the space spanned by the column vectors of  $\mathbf{H}_w$ , namely

$$\begin{aligned} \mathbf{U}_1' \mathbf{U}_1 &= \mathbf{D}^{-1/2} \mathbf{V}_1' \frac{\mathbf{H}_w' \mathbf{H}_w}{T} \mathbf{V}_1 \mathbf{D}^{-1/2} \\ &= \mathbf{D}^{-1/2} \mathbf{D} \mathbf{D}^{-1/2} \\ &= \mathbf{I}_{r_c p_T + 1}. \end{aligned}$$

Scaled matrix  $T^{-1/2} \mathbf{H}_w$  can now be written as  $T^{-1/2} \mathbf{H}_w = \mathbf{U}_1 \mathbf{D}^{1/2} \mathbf{V}_1'$ . Consider

$$\mathbf{D}^{-1/2} \mathbf{V}_1' \frac{\mathbf{H}_w' \boldsymbol{\varepsilon}_i}{T} = \mathbf{D}^{-1/2} \mathbf{V}_1' \mathbf{V}_1 \mathbf{D}^{1/2} \mathbf{U}_1' \boldsymbol{\varepsilon}_i^\diamond = \mathbf{U}_1' \boldsymbol{\varepsilon}_i^\diamond,$$

where we have used that  $\mathbf{V}_1' \mathbf{V}_1$  is an identity matrix since  $\mathbf{V}_1$  is unitary. Using now the submultiplicative property of matrix norms and (A.11) of Lemma A.2, we obtain

$$\begin{aligned} \|\mathbf{U}_1' \boldsymbol{\varepsilon}_i^\diamond\|_\infty &= \left\| \mathbf{D}^{-1/2} \mathbf{V}_1' \frac{\mathbf{H}_w' \boldsymbol{\varepsilon}_i}{T} \right\|_\infty \\ &\leq \left\| \mathbf{D}^{-1/2} \right\|_\infty \|\mathbf{V}_1'\|_\infty \left\| \frac{\mathbf{H}_w' \boldsymbol{\varepsilon}_i}{T} \right\|_\infty \\ &= O_p\left(T^{-1/2}\right), \end{aligned}$$

where  $\|\mathbf{D}^{-1/2}\|_\infty = O_p(1)$  since the diagonal elements of the diagonal matrix  $\mathbf{D}$  have positive probability limits, and  $\|\mathbf{V}_1'\|_\infty = O_p(1)$  since  $\mathbf{V}_1$  is unitary. This establishes that the individual elements of the vector  $\mathbf{U}_1' \boldsymbol{\varepsilon}_i^\diamond$  are (uniformly)  $O_p(T^{-1/2})$ . Consider next  $\mathbf{P}_h \boldsymbol{\varepsilon}_i^\diamond$ , which is an orthogonal projection of  $\boldsymbol{\varepsilon}_i^\diamond$  on the space spanned by the column vectors of  $\mathbf{H}_w$ . Since  $\mathbf{U}_1$  is an orthonormal basis of this space, we can write  $\mathbf{P}_h \boldsymbol{\varepsilon}_i^\diamond$  as the following linear combination of basis vectors,<sup>11</sup>

$$\mathbf{P}_h \boldsymbol{\varepsilon}_i^\diamond = \sum_{j=1}^{(r_c+1)p_T+1} \langle \boldsymbol{\varepsilon}_i^\diamond, \mathbf{u}_{1j} \rangle \mathbf{u}_{1j}, \quad (\text{A.35})$$

where  $\mathbf{u}_{1j}$ , for  $j = 1, 2, \dots, r_c p_T + 1$ , denote the individual columns of  $\mathbf{U}_1$ . But we have shown that  $|\langle \boldsymbol{\varepsilon}_i^\diamond, \mathbf{u}_{1j} \rangle| =$

---

<sup>11</sup>The column vectors in  $\mathbf{U}$  are orthogonal and therefore for any vector  $\mathbf{a} \in \text{Col}(\mathbf{U})$  we have  $\mathbf{a} = \sum_{j=1}^{r_c p_T + 1} \frac{\langle \mathbf{a}, \mathbf{u}_{1j} \rangle}{\langle \mathbf{u}_{1j}, \mathbf{u}_{1j} \rangle} \mathbf{u}_{1j}$ . But  $\langle \mathbf{u}_{1j}, \mathbf{u}_{1j} \rangle = 1$  since each of the column vectors contained in  $\mathbf{U}$  have unit length (orthonormality) and we obtain  $\mathbf{a} = \sum_{j=1}^{r_c p_T + 1} \langle \mathbf{a}, \mathbf{u}_{1j} \rangle \mathbf{u}_{1j}$ . (A.35) now follows by letting  $\mathbf{a} = \mathbf{P}_h \boldsymbol{\varepsilon}_i^\diamond$  and noting that  $\langle \mathbf{P}_h \boldsymbol{\varepsilon}_i^\diamond, \mathbf{u}_{1j} \rangle = \langle \boldsymbol{\varepsilon}_i^\diamond, \mathbf{u}_{1j} \rangle$  since  $\mathbf{P}_h \mathbf{u}_{1j} = \mathbf{u}_{1j}$ .

$O_p(T^{-1/2})$  and  $\|\mathbf{u}_{1j}\| = 1$  (orthonormality), and therefore

$$\|\mathbf{P}_h \boldsymbol{\varepsilon}_i^\diamond\| = O_p\left(\frac{pT}{\sqrt{T}}\right). \quad (\text{A.36})$$

Using (A.33) and (A.36) in (A.32) yields

$$\|\langle \mathbf{P}_h \boldsymbol{\xi}_{is}^\diamond, \mathbf{P}_h \boldsymbol{\varepsilon}_i^\diamond \rangle\| = O_p\left(\frac{pT}{\sqrt{T}}\right),$$

for  $s = 1, 2, \dots, 2k + 1$ , and using this result together with (A.31) in (A.30) we obtain

$$\|\langle \mathbf{M}_h \boldsymbol{\xi}_{is}^\diamond, \mathbf{M}_h \boldsymbol{\varepsilon}_i^\diamond \rangle\|_\infty \xrightarrow{p} 0,$$

as desired. This completes the proof of (A.17)

(A.18) and (A.19) can be established in a similar way by noting that Lemma A.2 implies  $\|T^{-1} \boldsymbol{\Xi}'_i \boldsymbol{\eta}_i\|_\infty \xrightarrow{p} 0$  and  $\|\langle \boldsymbol{\eta}_i^\diamond, T^{-1/2} \mathbf{H}_w \rangle\|_\infty = O_p(T^{-1/2})$  (required to establish (A.18)) and also  $\|T^{-1} \boldsymbol{\Xi}'_i \boldsymbol{\vartheta}_i\|_\infty \xrightarrow{p} 0$ ,  $\|\langle \boldsymbol{\vartheta}_i^\diamond, T^{-1/2} \mathbf{H}_w \rangle\|_\infty = O_p(T^{-1/2})$  (required for (A.19)). ■

**Proof of Lemma A.5.** Define

$$\boldsymbol{\varphi}_{iT} = \boldsymbol{\Sigma}_{i\xi}^{-1} \frac{\boldsymbol{\Xi}'_i \mathbf{M}_h \mathbf{F}}{T} \boldsymbol{\eta}_{\gamma i},$$

and consider the cross section average  $\bar{\boldsymbol{\varphi}}_T = N^{-1} \sum_{i=1}^N \boldsymbol{\varphi}_{iT}$ . Note that

$$E(\boldsymbol{\varphi}_{iT}) = \mathbf{0}_{2k_x+1 \times 1}, \quad (\text{A.37})$$

and

$$E(\boldsymbol{\varphi}_{iT} \boldsymbol{\varphi}_{jT}') = \mathbf{0}_{2k_x+1 \times 2k_x+1} \text{ for } i \neq j, i, j = 1, 2, \dots, N, \quad (\text{A.38})$$

since the unobserved common factors are serially uncorrelated and independently distributed of  $\boldsymbol{\eta}_{\gamma i}$ , and  $\boldsymbol{\eta}_{\gamma i}$  is independently distributed across  $i$ . Next, we show that the individual elements of  $E(\boldsymbol{\varphi}_{iT} \boldsymbol{\varphi}_{iT}')$  are bounded in  $N$ .  $\boldsymbol{\Sigma}_{i\xi}$  defined in Lemma A.3 is invertible under Assumption 7 and in particular  $\|\boldsymbol{\Sigma}_{i\xi}^{-1}\| < K < \infty$ . Using Cauchy-Schwarz inequality, we obtain

$$E\left[\left(\tilde{\xi}_{ist} f_{\ell t} \eta_{\gamma i \ell}\right)^2\right] \leq \sqrt{E\left(\tilde{\xi}_{ist}^4\right) E\left(f_{\ell t}^4 \eta_{\gamma i \ell}^4\right)} = O(1),$$

for  $s = 1, 2, \dots, 2k + 1$ , and  $\ell = 1, 2, \dots, m$ , where  $\tilde{\xi}_{ist}$  are the individual elements of  $\boldsymbol{\Xi}'_i \mathbf{M}_h$ ,  $\tilde{\xi}_{ist}$  has uniformly bounded 4-th moments under Assumption 7, and  $E\left(f_{\ell t}^4 \eta_{\gamma i \ell}^4\right) = E\left(f_{\ell t}^4\right) E\left(\eta_{\gamma i \ell}^4\right)$  is also uniformly bounded under Assumptions 2 and 3. It follows that there exists a constant  $K < \infty$ , which does not depend on  $N$  and such that

$$\|E(\boldsymbol{\varphi}_{iT} \boldsymbol{\varphi}_{iT}')\| < K. \quad (\text{A.39})$$

Using now (A.38)-(A.39), we obtain

$$\|Var(\bar{\boldsymbol{\varphi}}_T)\| = O(N^{-1}). \quad (\text{A.40})$$

(A.37) and (A.40) imply  $\bar{\boldsymbol{\varphi}}_T \xrightarrow{p} 0$ , as desired. ■

**Proof of Lemma A.6.** Denote the individual columns of  $\boldsymbol{\Xi}_i$  by  $\boldsymbol{\xi}_{is}$ ,  $s = 1, 2, \dots, 2k + 1$  and consider

$$\boldsymbol{\xi}_{is}' \bar{\mathbf{M}}_q \boldsymbol{\xi}_{is} - \boldsymbol{\xi}_{is}' \mathbf{M}_h \boldsymbol{\xi}_{is} = \|\bar{\mathbf{M}}_q \boldsymbol{\xi}_{is}\|^2 - \|\mathbf{M}_h \boldsymbol{\xi}_{is}\|^2, \quad (\text{A.41})$$

for  $s = 1, 2, \dots, 2k + 1$ . Hilbert projection theorem (see Rudin, 1987) implies

$$\|\bar{\mathbf{M}}_q \boldsymbol{\xi}_{is.}\|^2 \leq \|\boldsymbol{\xi}_{is.} - \boldsymbol{\alpha}_q\|^2,$$

for any vector  $\boldsymbol{\alpha}_q \in \text{Col}(\bar{\mathbf{Q}}_w)$ . Choose  $\boldsymbol{\alpha}_q = \mathbf{P}_h \boldsymbol{\xi}_{is.} - \bar{\mathbf{M}}_q \mathbf{P}_h \boldsymbol{\xi}_{is.}$ , where  $\mathbf{P}_h$  is orthogonal projector matrix onto  $\text{Col}(\bar{\mathbf{Q}}_w)$ , and note that  $\boldsymbol{\alpha}_q = (\mathbf{I}_{T-p_T} - \bar{\mathbf{M}}_q) \mathbf{P}_h \boldsymbol{\xi}_{is.} \in \text{Col}(\bar{\mathbf{Q}}_w)$ . Hence,

$$\begin{aligned} \|\bar{\mathbf{M}}_q \boldsymbol{\xi}_{is.}\|^2 &\leq \|\boldsymbol{\xi}_{is.} - \mathbf{P}_h \boldsymbol{\xi}_{is.} + \bar{\mathbf{M}}_q \mathbf{P}_h \boldsymbol{\xi}_{is.}\|^2 \\ &\leq \|\mathbf{M}_h \boldsymbol{\xi}_{is.} + \bar{\mathbf{M}}_q \mathbf{P}_h \boldsymbol{\xi}_{is.}\|^2 \\ &\leq \|\mathbf{M}_h \boldsymbol{\xi}_{is.}\|^2 + \|\bar{\mathbf{M}}_q \mathbf{P}_h \boldsymbol{\xi}_{is.}\|^2 + 2 \langle \mathbf{M}_h \boldsymbol{\xi}_{is.}, \bar{\mathbf{M}}_q \mathbf{P}_h \boldsymbol{\xi}_{is.} \rangle, \end{aligned} \quad (\text{A.42})$$

where we used  $\mathbf{M}_h = \mathbf{I}_{T-p_T} - \mathbf{P}_h$  to obtain the second inequality, and we used  $\|\mathbf{a} + \mathbf{b}\|^2 = \|\mathbf{a}\|^2 + \|\mathbf{b}\|^2 + 2 \langle \mathbf{a}, \mathbf{b} \rangle$ , for any vectors  $\mathbf{a}$  and  $\mathbf{b}$ , to obtain the third inequality. Similarly, we obtain the following upper bound on  $\|\mathbf{M}_h \boldsymbol{\xi}_{is.}\|^2$ ,

$$\begin{aligned} \|\mathbf{M}_h \boldsymbol{\xi}_{is.}\|^2 &\leq \|\boldsymbol{\xi}_{is.} - \bar{\mathbf{P}}_q \boldsymbol{\xi}_{is.} + \mathbf{M}_h \bar{\mathbf{P}}_q \boldsymbol{\xi}_{is.}\|^2 \\ &\leq \|\bar{\mathbf{M}}_q \boldsymbol{\xi}_{is.} + \mathbf{M}_h \bar{\mathbf{P}}_q \boldsymbol{\xi}_{is.}\|^2 \\ &\leq \|\bar{\mathbf{M}}_q \boldsymbol{\xi}_{is.}\|^2 + \|\mathbf{M}_h \bar{\mathbf{P}}_q \boldsymbol{\xi}_{is.}\|^2 + 2 \langle \bar{\mathbf{M}}_q \boldsymbol{\xi}_{is.}, \mathbf{M}_h \bar{\mathbf{P}}_q \boldsymbol{\xi}_{is.} \rangle \end{aligned} \quad (\text{A.43})$$

Using (A.42) and (A.43) in (A.41) yields the following lower and upper bounds,

$$\epsilon_{1,NT} \leq \|\bar{\mathbf{M}}_q \boldsymbol{\xi}_{is.}\|^2 - \|\mathbf{M}_h \boldsymbol{\xi}_{is.}\|^2 \leq \epsilon_{2,NT}, \quad (\text{A.44})$$

where

$$\epsilon_{1,NT} = \|\mathbf{M}_h \bar{\mathbf{P}}_q \boldsymbol{\xi}_{is.}\|^2 + 2 \langle \bar{\mathbf{M}}_q \boldsymbol{\xi}_{is.}, \mathbf{M}_h \bar{\mathbf{P}}_q \boldsymbol{\xi}_{is.} \rangle, \quad (\text{A.45})$$

and

$$\epsilon_{2,NT} = \|\bar{\mathbf{M}}_q \mathbf{P}_h \boldsymbol{\xi}_{is.}\|^2 + 2 \langle \mathbf{M}_h \boldsymbol{\xi}_{is.}, \bar{\mathbf{M}}_q \mathbf{P}_h \boldsymbol{\xi}_{is.} \rangle. \quad (\text{A.46})$$

Note that  $\bar{\mathbf{P}}_q \boldsymbol{\xi}_{is.}$  belongs to  $\text{Col}(\bar{\mathbf{Q}}_w)$  and  $\|\bar{\mathbf{P}}_q \boldsymbol{\xi}_{is.}\| \leq \|\boldsymbol{\xi}_{is.}\| = O_p(\sqrt{T-p_T})$  since the individual elements of  $\boldsymbol{\xi}_{is.}$  are uniformly  $O_p(1)$ . Also,  $\bar{\mathbf{Q}}_w = \mathbf{H}_w + \bar{\mathbf{V}}_w$ , where elements of  $\bar{\mathbf{V}}_w$  are uniformly  $O_p(N^{-1/2})$ , whereas the elements of  $\mathbf{H}_w$  are  $O_p(1)$ . Using Lemma A.1 (by setting  $\mathbf{A} = \mathbf{H}_w + \bar{\mathbf{V}}_w$ ,  $\mathbf{B} = -\bar{\mathbf{V}}_w$  and  $\boldsymbol{\alpha}_{A,1} = \bar{\mathbf{P}}_q \boldsymbol{\xi}_{is.}$ ), we obtain

$$\|\mathbf{M}_h \bar{\mathbf{P}}_q \boldsymbol{\xi}_{is.}\| = O_p\left(\frac{p_T \sqrt{T-p_T}}{\sqrt{N}}\right). \quad (\text{A.47})$$

Similarly, Lemma A.1 can be used again (by setting  $\mathbf{A} = \mathbf{H}_w$ ,  $\mathbf{B} = \bar{\mathbf{V}}_w$  and  $\boldsymbol{\alpha}_{A,1} = \mathbf{P}_h \boldsymbol{\xi}_{is.}$ ) to show that

$$\|\bar{\mathbf{M}}_q \mathbf{P}_h \boldsymbol{\xi}_{is.}\| = O_p\left(\frac{p_T \sqrt{T-p_T}}{\sqrt{N}}\right). \quad (\text{A.48})$$

Now consider the inner product on the right side of (A.45). Using Cauchy-Schwarz inequality, we have

$$\begin{aligned} |\langle \bar{\mathbf{M}}_q \boldsymbol{\xi}_{is.}, \mathbf{M}_h \bar{\mathbf{P}}_q \boldsymbol{\xi}_{is.} \rangle| &\leq \|\bar{\mathbf{M}}_q \boldsymbol{\xi}_{is.}\| \|\mathbf{M}_h \bar{\mathbf{P}}_q \boldsymbol{\xi}_{is.}\|, \\ &= O_p\left(\frac{p_T (T-p_T)}{\sqrt{N}}\right) \end{aligned} \quad (\text{A.49})$$

where  $\|\bar{\mathbf{M}}_q \boldsymbol{\xi}_{is.}\| \leq \|\boldsymbol{\xi}_{is.}\| = O_p(\sqrt{T-p_T})$ , and  $\|\mathbf{M}_h \bar{\mathbf{P}}_q \boldsymbol{\xi}_{is.}\| = O_p(p_T N^{-1/2} \sqrt{T-p_T})$  by (A.47). Similarly,

using  $\|\mathbf{M}_h \boldsymbol{\xi}_{is}\| \leq \|\boldsymbol{\xi}_{is}\| = O_p(\sqrt{T - p_T})$ , (A.48) and the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned} |\langle \mathbf{M}_h \boldsymbol{\xi}_{is}, \bar{\mathbf{M}}_q \mathbf{P}_h \boldsymbol{\xi}_{is} \rangle| &\leq \|\mathbf{M}_h \boldsymbol{\xi}_{is}\| \|\bar{\mathbf{M}}_q \mathbf{P}_h \boldsymbol{\xi}_{is}\| \\ &= O_p\left(\frac{p_T(T - p_T)}{\sqrt{N}}\right) \end{aligned} \quad (\text{A.50})$$

Using (A.47)-(A.50) in (A.45) and (A.46) we obtain

$$\epsilon_{\ell, NT} = O_p\left(\frac{p_T^2(T - p_T)^2}{N}\right) + O_p\left(\frac{p_T(T - p_T)}{\sqrt{N}}\right), \text{ for } \ell = 1, 2;$$

and using this result in (A.44) yields

$$\begin{aligned} \sqrt{N} \left( \left\| \frac{\bar{\mathbf{M}}_q \boldsymbol{\xi}_{is}}{T} \right\|^2 - \left\| \frac{\mathbf{M}_h \boldsymbol{\xi}_{is}}{T} \right\|^2 \right) &= O_p\left(\frac{p_T^2(T - p_T)}{T^2 \sqrt{N}}\right) + O_p\left(\frac{p_T(T - p_T)}{T^2}\right), \\ &\xrightarrow{p} 0, \end{aligned}$$

for  $s = 1, 2, \dots, 2k + 1$ , as  $(N, T, p_T) \rightarrow \infty$  such that  $p_T^2/T \rightarrow 0$ . This establishes that the diagonal elements of

$$\sqrt{N} \frac{\boldsymbol{\Xi}'_i \bar{\mathbf{M}}_q \boldsymbol{\Xi}_i}{T} - \sqrt{N} \frac{\boldsymbol{\Xi}'_i \mathbf{M}_h \boldsymbol{\Xi}_i}{T}$$

tend to 0 in probability uniformly in  $i$ .

Now consider the off-diagonal elements. Convergence of individual terms

$$\sqrt{N} \frac{\boldsymbol{\xi}'_{is} \bar{\mathbf{M}}_q \boldsymbol{\xi}_{i\ell}}{T} - \sqrt{N} \frac{\boldsymbol{\xi}'_{is} \mathbf{M}_h \boldsymbol{\xi}_{i\ell}}{T}, \text{ for } s \neq \ell, s, \ell = 1, 2, \dots, k + 1,$$

can be established following the same arguments as above, but using (A.8) instead of (A.7) of Lemma A.1. This completes the proof of (A.21). (A.22)-(A.25) can be established in the same way. ■

**Proof of Lemma A.7.** Using the identity  $\mathbf{M}_h = \mathbf{I}_{T-p_T} - \mathbf{P}_h$ , where  $\mathbf{P}_h$  is orthogonal projection matrix that projects onto  $Col(\mathbf{H}_w)$ , we write the expression on the left side of (A.26) as:

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{\boldsymbol{\Xi}'_i \mathbf{M}_h \boldsymbol{\epsilon}_i}{T} = \frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{\boldsymbol{\Xi}'_i \boldsymbol{\epsilon}_i}{T} - \frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{\boldsymbol{\Xi}'_i \mathbf{P}_h \boldsymbol{\epsilon}_i}{T}. \quad (\text{A.51})$$

First we establish convergence of the first term on the right side of (A.51). Let  $T_N = T(N)$  and  $p_N = p_T[T(N)]$  be any non-decreasing integer-valued functions of  $N$  such that  $\lim_{N \rightarrow \infty} T_N = \infty$  and  $\lim_{N \rightarrow \infty} p_N^2/T = 0$ . The first term on the right side of (A.51) can be written as

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{\boldsymbol{\Xi}'_i \boldsymbol{\epsilon}_i}{T_N} = \sum_{t=p_T+1}^{T_N} \boldsymbol{\kappa}_{Nt},$$

where

$$\boldsymbol{\kappa}_{Nt} = \frac{1}{T_N \sqrt{N}} \sum_{i=1}^N \boldsymbol{\xi}_{it} \boldsymbol{\epsilon}_{it}.$$

Let  $\{\{c_{Nt}\}_{t=-\infty}^{\infty}\}_{N=1}^{\infty}$  be two-dimensional array of constants and set  $c_{Nt} = \frac{1}{T_N}$  for all  $t \in \mathbb{Z}$  and  $N \in \mathbb{N}$ .  $\boldsymbol{\xi}_{it}$  and  $\boldsymbol{\epsilon}_{jt}$  are independently distributed for any  $i, j$  and  $t$ , and we have:  $E(\boldsymbol{\kappa}_{Nt}) = 0$ , and the elements of

covariance matrix of  $\boldsymbol{\kappa}_{Nt}/c_{Nt}$  are bounded, in particular

$$\begin{aligned} \left\| \text{Var} \left( \frac{\boldsymbol{\kappa}_{Nt}}{c_{Nt}} \right) \right\| &= \left\| E \left( \frac{\boldsymbol{\kappa}_{Nt} \boldsymbol{\kappa}_{Nt}'}{c_{Nt}^2} \right) \right\|, \\ &= \left\| \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N E \left( \boldsymbol{\xi}_{it} \boldsymbol{\xi}_{jt}' \varepsilon_{it} \varepsilon_{jt} \right) \right\|, \\ &= \left\| \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N [E \left( \boldsymbol{\xi}_{it} \boldsymbol{\xi}_{jt}' \right) E \left( \varepsilon_{it} \varepsilon_{jt} \right)] \right\|. \end{aligned}$$

Noting that  $E \left( \boldsymbol{\xi}_{it} \boldsymbol{\xi}_{jt}' \right)$  is bounded in  $i, j$  and  $t$ , and  $E \left( \varepsilon_t \varepsilon_t' \right) = \mathbf{R} \mathbf{R}'$  under Assumption 1, we obtain

$$\begin{aligned} \left\| \text{Var} \left( \frac{\boldsymbol{\kappa}_{Nt}}{c_{Nt}} \right) \right\| &\leq \frac{K}{N} \left\| \sum_{i=1}^N \sum_{j=1}^N E \left( \varepsilon_{it} \varepsilon_{jt} \right) \right\|, \\ &\leq \frac{K}{N} \left\| \boldsymbol{\tau}' E \left( \varepsilon_t \varepsilon_t' \right) \boldsymbol{\tau} \right\|, \\ &\leq \frac{K}{N} \left\| \boldsymbol{\tau}'_N \right\| \left\| \mathbf{R} \right\| \left\| \mathbf{R}' \right\| \left\| \boldsymbol{\tau}_N \right\|. \end{aligned}$$

But  $\left\| \boldsymbol{\tau}'_N \right\| = \left\| \boldsymbol{\tau}_N \right\| = \sqrt{N}$  and  $\left\| \mathbf{R} \right\| \leq \sqrt{\left\| \mathbf{R} \right\|_1 \left\| \mathbf{R} \right\|_\infty} < K$ , where  $\left\| \mathbf{R} \right\|_1$  and  $\left\| \mathbf{R} \right\|_\infty$  are postulated to be bounded by Assumption 1, and therefore

$$\left\| \text{Var} \left( \frac{\boldsymbol{\kappa}_{Nt}}{c_{Nt}} \right) \right\| = O(1). \quad (\text{A.52})$$

(A.52) implies uniform integrability of  $\{\boldsymbol{\kappa}_{Nt}/c_{Nt}\}$  and the array  $\boldsymbol{\kappa}_{Nt}$  is uniformly integrable  $L_1$ -mixingale array with respect to the constant array  $c_{Nt}$ . Using a mixingale weak law yields (Davidson, 1994, Theorem 19.11)

$$\sum_{t=p_T+1}^{T_N} \boldsymbol{\kappa}_{Nt} = \frac{1}{T_N \sqrt{N}} \sum_{t=p_T+1}^{T_N} \sum_{i=1}^N \boldsymbol{\xi}_{it} \varepsilon_{it} \xrightarrow{L_1} \mathbf{0}_{2k_x+1 \times 1}.$$

Convergence in  $L_1$  norm implies convergence in probability. This establishes

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{\boldsymbol{\Xi}'_i \boldsymbol{\varepsilon}_i}{T} \xrightarrow{p} \mathbf{0}_{2k_x+1 \times 1}, \quad (\text{A.53})$$

as  $(N, T, p_T) \xrightarrow{j} \infty$  and  $p_T^2/T \rightarrow 0$ .

Next consider the second term on the right hand side of (A.51), and note that

$$\begin{aligned} \frac{\boldsymbol{\Xi}'_i \mathbf{P}_h \boldsymbol{\varepsilon}_i}{T} &= \frac{1}{\sqrt{T}} \frac{\boldsymbol{\Xi}'_i \mathbf{H}_w}{T} \left( \frac{\mathbf{H}'_w \mathbf{H}_w}{T} \right)^+ \frac{\mathbf{H}'_w \boldsymbol{\varepsilon}_i}{\sqrt{T}}, \\ &= \frac{1}{\sqrt{T}} \mathbf{G}'_{iT} \boldsymbol{\vartheta}_{\varepsilon i}, \end{aligned}$$

where

$$\mathbf{G}'_{iT} = \frac{\boldsymbol{\Xi}'_i \mathbf{H}_w}{T} \left( \frac{\mathbf{H}'_w \mathbf{H}_w}{T} \right)^+,$$

and

$$\boldsymbol{\vartheta}_{\varepsilon i} = \frac{\mathbf{H}'_w \boldsymbol{\varepsilon}_i}{\sqrt{T}}.$$

Define also

$$\mathbf{G}'_i = \mathbf{\Theta}_{\xi i} \mathbf{\Theta}_{hh}^+,$$

in which  $\mathbf{\Theta}_{\xi i} = E(\xi_{it} \tilde{\mathbf{h}}'_{wt})$ ,  $\tilde{\mathbf{h}}'_{wt} = (1, \mathbf{h}'_{wt}, \mathbf{h}'_{w,t-1}, \dots, \mathbf{h}'_{w,t-p_T})$  denotes the individual rows of  $\mathbf{H}_w$ , is  $2k_x + 1 \times (k+1)p_T + 1$  dimensional matrix, and  $\mathbf{\Theta}_{hh} = E(\tilde{\mathbf{h}}_{wt} \tilde{\mathbf{h}}'_{wt})$  is  $(k+1)p_T + 1 \times (k+1)p_T + 1$  matrix. Elements of  $\mathbf{\Theta}_{\xi i}$  and  $\mathbf{\Theta}_{hh}$  are uniformly bounded and in particular

$$\|\mathbf{\Theta}_{hh}^+\|_\infty = O(1), \|\mathbf{\Theta}_{hh}^+\|_1 = O(1), \|\mathbf{\Theta}_{\xi i}\|_\infty = O(1) \text{ and } \|\mathbf{\Theta}_{\xi i}\|_1 = O(1), \quad (\text{A.54})$$

because  $\sum_{\ell=0}^\infty |E(\xi_{ist} h_{w,r_2 t-\ell})| < K$  and  $\sum_{\ell=0}^\infty |E(h_{w,r_1 t} h_{w,r_2 t-\ell})| < K$  for any  $r_1, r_2 = 1, 2, \dots, k+1$  and  $s = 1, 2, \dots, k+1$ , where  $h_{w,r_1 t}$  for  $r_1 = 1, 2, \dots, k+1$  denotes individual elements of  $\mathbf{h}_{wt} = \mathbf{\Psi}_w(L) \mathbf{f}_t + \mathbf{c}_{zw}$  and  $\xi_{ist}$  for  $s = 1, 2, \dots, k+1$  denotes individual elements of  $\xi_{it}$ . Using these notations, we can now write the second term on the right side of (A.51) as

$$\begin{aligned} \frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{\mathbf{\Xi}'_i \mathbf{P}_h \boldsymbol{\varepsilon}_i}{T} &= \sqrt{\frac{N}{T}} \cdot \frac{1}{N} \sum_{i=1}^N \mathbf{G}'_{iT} \boldsymbol{\vartheta}_{\varepsilon i} \\ &= \sqrt{\frac{N}{T}} \left( \frac{1}{N} \sum_{i=1}^N \mathbf{G}'_i \boldsymbol{\vartheta}_{\varepsilon i} + \frac{1}{N} \sum_{i=1}^N (\mathbf{G}'_{iT} - \mathbf{G}'_i) \boldsymbol{\vartheta}_{\varepsilon i} \right) \end{aligned} \quad (\text{A.55})$$

Consider the first term inside the brackets on the right side of (A.55), and note that

$$E \left( \frac{1}{N} \sum_{i=1}^N \mathbf{G}'_i \boldsymbol{\vartheta}_{\varepsilon i} \right) \left( \frac{1}{N} \sum_{i=1}^N \mathbf{G}'_i \boldsymbol{\vartheta}_{\varepsilon i} \right)' = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \mathbf{G}'_i E(\boldsymbol{\vartheta}_{\varepsilon i} \boldsymbol{\vartheta}'_{\varepsilon j}) \mathbf{G}_j. \quad (\text{A.56})$$

Since  $\boldsymbol{\varepsilon}_i$  is independently distributed of  $\tilde{\mathbf{h}}'_{wt}$  and the stochastic processes in  $\tilde{\mathbf{h}}'_{wt}$  are covariance stationary we also have

$$E(\boldsymbol{\vartheta}_{\varepsilon i} \boldsymbol{\vartheta}'_{\varepsilon j}) = \frac{1}{T} E(\mathbf{H}'_w \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}'_j \mathbf{H}_w) = \sigma_{ij} \mathbf{\Theta}_{hh}, \quad (\text{A.57})$$

where  $\sigma_{ij} = E(\varepsilon_{it} \varepsilon_{jt})$ . Using (A.57) in (A.56) and applying the submultiplicative property of matrix norm yields

$$\begin{aligned} \left\| E \left( \frac{1}{N} \sum_{i=1}^N \mathbf{G}'_i \boldsymbol{\vartheta}_{\varepsilon i} \right) \left( \frac{1}{N} \sum_{i=1}^N \mathbf{G}'_i \boldsymbol{\vartheta}_{\varepsilon i} \right)' \right\|_\infty &= \left\| \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \sigma_{ij} \mathbf{G}'_i \mathbf{\Theta}_{hh} \mathbf{G}_j \right\|_\infty \\ &\leq \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N |\sigma_{ij}| \|\mathbf{G}'_i\|_\infty \|\mathbf{\Theta}_{hh}\|_\infty \|\mathbf{G}_j\|_\infty, \end{aligned}$$

where  $\|\mathbf{\Theta}_{hh}\|_\infty = O(1)$ ,  $\|\mathbf{G}'_i\|_\infty = \|\mathbf{\Theta}_{\xi i} \mathbf{\Theta}_{hh}\|_\infty \leq \|\mathbf{\Theta}_{\xi i}\|_\infty \|\mathbf{\Theta}_{hh}\|_\infty = O(1)$ , and  $\|\mathbf{G}_j\|_\infty = \|(\mathbf{\Theta}_{\xi j} \mathbf{\Theta}_{hh})'\|_\infty = \|\mathbf{\Theta}_{\xi j} \mathbf{\Theta}_{hh}\|_1 \leq \|\mathbf{\Theta}_{\xi j}\|_1 \|\mathbf{\Theta}_{hh}\|_1 = O(1)$ , see (A.54). Using these results and noting that  $N^{-1} \sum_{i=1}^N \sum_{j=1}^N |\sigma_{ij}| = O(1)$  under Assumption 1, we obtain

$$\begin{aligned} \left\| E \left( \frac{1}{N} \sum_{i=1}^N \mathbf{G}'_i \boldsymbol{\vartheta}_{\varepsilon i} \right) \left( \frac{1}{N} \sum_{i=1}^N \mathbf{G}'_i \boldsymbol{\vartheta}_{\varepsilon i} \right)' \right\|_\infty &\leq \frac{K}{N^2} \sum_{i=1}^N \sum_{j=1}^N |\sigma_{ij}| \\ &\leq \frac{K}{N}, \end{aligned} \quad (\text{A.58})$$

which in turn implies that

$$\sqrt{\frac{N}{T}} \cdot \frac{1}{N} \sum_{i=1}^N \mathbf{G}'_i \boldsymbol{\vartheta}_{\varepsilon i} \xrightarrow{p} \mathbf{0}_{2k_x+1 \times 1}, \quad (\text{A.59})$$

as  $(N, T, p_T) \xrightarrow{j} \infty$  such that  $N/T \rightarrow \varkappa_1$ , for some  $0 < \varkappa_1 < \infty$ .

Now consider the second term inside the brackets on the right side of (A.55). Using submultiplicative property of matrix norms, we have

$$\left\| \frac{1}{N} \sum_{i=1}^N (\mathbf{G}'_{iT} - \mathbf{G}'_i) \boldsymbol{\vartheta}_{\varepsilon i} \right\|_{\infty} \leq \frac{1}{N} \sum_{i=1}^N \|\mathbf{G}'_{iT} - \mathbf{G}'_i\|_{\infty} \|\boldsymbol{\vartheta}_{\varepsilon i}\|_{\infty}. \quad (\text{A.60})$$

Note that  $\boldsymbol{\vartheta}_{\varepsilon i}$  has zero mean and  $\text{Var}(\boldsymbol{\vartheta}_{\varepsilon i}) = E(\boldsymbol{\vartheta}_{\varepsilon i} \boldsymbol{\vartheta}'_{\varepsilon j}) = \sigma_{ij} \boldsymbol{\Theta}_{hh}$ , see (A.57), where  $\sigma_{ij}$  and the elements of  $\boldsymbol{\Theta}_{hh}$  are uniformly bounded. It therefore follows that

$$\|\boldsymbol{\vartheta}_{\varepsilon i}\|_{\infty} = O_p(1) \text{ uniformly in } i \text{ and } p_T. \quad (\text{A.61})$$

Consider now the term  $\sqrt{T} \|\mathbf{G}'_{iT} - \mathbf{G}'_i\|_{\infty}$ , and first note that

$$\begin{aligned} \mathbf{G}'_{iT} - \mathbf{G}'_i &= \frac{\boldsymbol{\Xi}'_i \mathbf{H}_w}{T} \left( \frac{\mathbf{H}'_w \mathbf{H}_w}{T} \right)^+ - \boldsymbol{\Theta}_{\xi i} \boldsymbol{\Theta}_{hh}^+ \\ &= \left[ \frac{\boldsymbol{\Xi}'_i \mathbf{H}_w}{T} - \boldsymbol{\Theta}_{\xi i} \right] \left[ \left( \frac{\mathbf{H}'_w \mathbf{H}_w}{T} \right)^+ - \boldsymbol{\Theta}_{hh}^+ \right] + \left[ \frac{\boldsymbol{\Xi}'_i \mathbf{H}_w}{T} - \boldsymbol{\Theta}_{\xi i} \right] \boldsymbol{\Theta}_{hh}^+ \\ &\quad + \boldsymbol{\Theta}_{\xi i} \left[ \left( \frac{\mathbf{H}'_w \mathbf{H}_w}{T} \right)^+ - \boldsymbol{\Theta}_{hh}^+ \right]. \end{aligned}$$

Hence

$$\begin{aligned} \|\mathbf{G}'_{iT} - \mathbf{G}'_i\|_{\infty} &\leq \left\| \left( \frac{\boldsymbol{\Xi}'_i \mathbf{H}_w}{T} - \boldsymbol{\Theta}_{\xi i} \right) \right\|_{\infty} \left\| \left( \frac{\mathbf{H}'_w \mathbf{H}_w}{T} \right)^+ - \boldsymbol{\Theta}_{hh}^+ \right\|_{\infty} + \left\| \frac{\boldsymbol{\Xi}'_i \mathbf{H}_w}{T} - \boldsymbol{\Theta}_{\xi i} \right\|_{\infty} \|\boldsymbol{\Theta}_{hh}^+\|_{\infty} \\ &\quad + \|\boldsymbol{\Theta}_{\xi i}\|_{\infty} \left\| \left[ \left( \frac{\mathbf{H}'_w \mathbf{H}_w}{T} \right)^+ - \boldsymbol{\Theta}_{hh}^+ \right] \right\|_{\infty} \end{aligned} \quad (\text{A.62})$$

Individual elements of  $\boldsymbol{\Xi}'_i \mathbf{H}_w / T - \boldsymbol{\Theta}_{\xi i}$  can be written as  $\sum_{t=p_T+1}^T \xi_{i,r,t} \tilde{h}'_{w,s,t} - E(\xi_{i,r,t} \tilde{h}'_{w,s,t})$ , for  $r = 1, 2, \dots, k+1$  and  $s = 1, 2, \dots, (k+1)p_T + 1$ , where  $\xi_{i,r,t}$  and  $\tilde{h}'_{w,s,t}$  are the elements of  $\boldsymbol{\xi}_{it}$  and  $\tilde{\mathbf{h}}_{wt}$ . The stochastic processes  $\xi_{i,r,t}$  and  $\tilde{h}'_{w,s,t}$  are covariance stationary with absolute summable autocovariances and we have  $\sum_{t=p_T+1}^T \xi_{i,r,t} \tilde{h}'_{w,s,t} - E(\xi_{i,r,t} \tilde{h}'_{w,s,t}) = O_p(T^{-1/2})$  uniformly in  $i$  and  $p_T$ . This implies

$$\left\| \left( \frac{\boldsymbol{\Xi}'_i \mathbf{H}_w}{T} - \boldsymbol{\Theta}_{\xi i} \right) \right\|_{\infty} = O_p\left(\frac{p_T}{\sqrt{T}}\right) \text{ uniformly in } i. \quad (\text{A.63})$$

Lemmas A.7 and A.8 of Chudik and Pesaran (2013b) establish that in the full column rank case where  $\text{rank}(\mathbf{C}) = m$  and  $k+1 = m$ , we have

$$\left\| \left( \frac{\mathbf{H}'_w \mathbf{H}_w}{T} \right)^{-1} - \boldsymbol{\Theta}_{hh}^{-1} \right\|_{\infty} = O_p\left(\frac{p_T}{\sqrt{T}}\right),$$

where  $\boldsymbol{\Theta}_{hh} = E(\tilde{\mathbf{h}}_{wt} \tilde{\mathbf{h}}'_{wt})$  is  $(k+1)p_T + 1 \times (k+1)p_T + 1$  nonsingular matrix (in the full column rank

case with  $k + 1 = m$ ). Using generalized inverse instead of inverse, the diagonalization of  $\mathbf{H}_w' \mathbf{H}_w / T$  in (A.34) and similar arguments as in Lemmas A.7 and A.8 of Chudik and Pesaran (2013b), the same result can be established for the more general case when  $\mathbf{C}$  does not necessarily have full column rank or when  $\text{rank}(\mathbf{C}) = m$  but  $k + 1 \geq m$ , namely:

$$\left\| \left( \frac{\mathbf{H}_w' \mathbf{H}_w}{T} \right)^+ - \boldsymbol{\Theta}_{hh}^+ \right\|_\infty = O_p \left( \frac{p_T}{\sqrt{T}} \right) \quad (\text{A.64})$$

Using (A.54) and (A.63)-(A.64) in (A.62), we obtain

$$\|\mathbf{G}_{iT}' - \mathbf{G}_i'\|_\infty = O_p \left( \frac{p_T}{\sqrt{T}} \right), \text{ uniformly in } i. \quad (\text{A.65})$$

Using now (A.61) together with (A.65) in (A.60) yield

$$\frac{1}{N} \sum_{i=1}^N (\mathbf{G}_{iT}' - \mathbf{G}_i') \boldsymbol{\vartheta}_{\varepsilon i} \xrightarrow{p} \mathbf{0}_{2k_x+1 \times 1}, \quad (\text{A.66})$$

as  $(N, T, p_T) \xrightarrow{j} \infty$ , and  $p_T^2/T \rightarrow 0$ . Finally, using (A.59) and (A.66) in (A.55), we obtain

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{\boldsymbol{\Xi}_i' \mathbf{P}_h \boldsymbol{\varepsilon}_i}{T} \xrightarrow{p} \mathbf{0}_{2k_x+1 \times 1}, \quad (\text{A.67})$$

when  $(N, T, p_T) \xrightarrow{j} \infty$  such that  $N/T \rightarrow \varkappa$ , for some  $0 < \varkappa < \infty$ , and  $p_T^2/T \rightarrow 0$ . This completes the proof.  $\blacksquare$

## A.4 Proofs of Theorems and Propositions

**Proof of Theorem 1.** Equation (24), for  $t = p_T + 1, p_T + 2, \dots, T$ , can be written as (see (A.2))

$$\mathbf{y}_i = \boldsymbol{\Xi}_i \boldsymbol{\pi}_i + \bar{\mathbf{Q}}_w \mathbf{d}_i + \boldsymbol{\varepsilon}_i + \boldsymbol{\eta}_i + \boldsymbol{\vartheta}_i, \quad (\text{A.68})$$

where  $\mathbf{d}_i = (c_{yi}^*, \boldsymbol{\delta}_{i0}', \boldsymbol{\delta}_{i1}', \dots, \boldsymbol{\delta}_{ip_T}')'$ ,  $\boldsymbol{\varepsilon}_i = (\varepsilon_{i,p_T+1}, \varepsilon_{i,p_T+2}, \dots, \varepsilon_{iT})'$ ,  $\boldsymbol{\eta}_i$  is  $T - p_T \times 1$  vector with its elements given by  $\sum_{\ell=p_T+1}^\infty \boldsymbol{\delta}_{i\ell}' \bar{\mathbf{z}}_{w,t-\ell}$ , for  $t = p_T + 1, p_T + 2, \dots, T$ , and  $\boldsymbol{\vartheta}_i$  is  $T - p_T \times 1$  vector defined in (A.3) with its elements uniformly bounded by  $O_p(N^{-1/2})$ . Substituting (A.68) into the definition of  $\hat{\boldsymbol{\pi}}_i$  in (26) and noting that  $(\boldsymbol{\Xi}_i' \bar{\mathbf{M}}_q \boldsymbol{\Xi}_i')^{-1} \boldsymbol{\Xi}_i' \bar{\mathbf{M}}_q \boldsymbol{\Xi}_i \boldsymbol{\pi}_i = \boldsymbol{\pi}_i$ , we obtain

$$\hat{\boldsymbol{\pi}}_i - \boldsymbol{\pi}_i = (\boldsymbol{\Xi}_i' \bar{\mathbf{M}}_q \boldsymbol{\Xi}_i')^{-1} \boldsymbol{\Xi}_i' \bar{\mathbf{M}}_q (\bar{\mathbf{Q}}_w \mathbf{d}_i + \boldsymbol{\varepsilon}_i + \boldsymbol{\eta}_i + \boldsymbol{\vartheta}_i). \quad (\text{A.69})$$

Note that  $\bar{\mathbf{M}}_q \bar{\mathbf{Q}}_w = \bar{\mathbf{Q}}_w - \bar{\mathbf{Q}}_w (\bar{\mathbf{Q}}_w' \bar{\mathbf{Q}}_w)^+ \bar{\mathbf{Q}}_w' \bar{\mathbf{Q}}_w = \bar{\mathbf{Q}}_w - \bar{\mathbf{Q}}_w = \mathbf{0}_{T-p_T \times (k+1)p_T+1}$  and (A.69) reduces to

$$\hat{\boldsymbol{\pi}}_i - \boldsymbol{\pi}_i = \left( \frac{\boldsymbol{\Xi}_i' \bar{\mathbf{M}}_q \boldsymbol{\Xi}_i'}{T} \right)^{-1} \frac{\boldsymbol{\Xi}_i' \bar{\mathbf{M}}_q}{T} (\boldsymbol{\varepsilon}_i + \boldsymbol{\eta}_i + \boldsymbol{\vartheta}_i) \quad (\text{A.70})$$

Consider the asymptotics  $(N, T, p_T) \xrightarrow{j} \infty$ , such that  $p_T^2/T \rightarrow \varkappa$ , for some constant  $0 < \varkappa < \infty$ . (A.12) of Lemma A.3 and (A.21) of Lemma A.6 show that  $T^{-1} \boldsymbol{\Xi}_i' \bar{\mathbf{M}}_q \boldsymbol{\Xi}_i'$  converges in probability to a full rank matrix and therefore

$$\left( \frac{\boldsymbol{\Xi}_i' \bar{\mathbf{M}}_q \boldsymbol{\Xi}_i'}{T} \right)^{-1} = O_p(1). \quad (\text{A.71})$$



Moreover, Lemmas A.4 and A.6 establish

$$\frac{\Xi'_i \bar{\mathbf{M}}_q \boldsymbol{\varepsilon}_i}{T} \xrightarrow{p} \mathbf{0}_{2k_x+1 \times 1}, \quad \frac{\Xi'_i \bar{\mathbf{M}}_q \boldsymbol{\eta}_i}{T} \xrightarrow{p} \mathbf{0}_{2k_x+1 \times 1}, \quad \text{and} \quad \frac{\Xi'_i \bar{\mathbf{M}}_q \boldsymbol{\vartheta}_i}{T} \xrightarrow{p} \mathbf{0}_{2k_x+1 \times 1}. \quad (\text{A.72})$$

Using (A.71)-(A.72) in (A.70) establish (28), as desired.  $\blacksquare$

**Proof of Theorem 2.** First suppose that the rank condition stated in Assumption 6 holds and consider the asymptotics  $(N, T, p_T) \xrightarrow{j} \infty$ , such that  $p_T^3/T \rightarrow \varkappa$ , for some constant  $0 < \varkappa < \infty$ . Using Theorem 1 and the definition of the mean group estimator  $\hat{\boldsymbol{\pi}}_{MG}$  in (27), we have

$$\hat{\boldsymbol{\pi}}_{MG} - \frac{1}{N} \sum_{i=1}^N \boldsymbol{\pi}_i \xrightarrow{p} \mathbf{0}_{2k_x+1 \times 1}. \quad (\text{A.73})$$

Assumption 4 postulates that  $\boldsymbol{\pi}_i = \boldsymbol{\pi} + \mathbf{v}_{\pi i}$ , where  $\mathbf{v}_{\pi i} \sim IID \left( \mathbf{0}_{2k_x+1 \times 1}, \boldsymbol{\Omega}_\pi \right)$  and the norms of  $\boldsymbol{\pi}$  and  $\boldsymbol{\Omega}_\pi$  are bounded. It follows that  $\left\| Var \left( N^{-1} \sum_{i=1}^N \mathbf{v}_{\pi i} \right) \right\| = \|\boldsymbol{\Omega}_\pi/N\| \rightarrow 0$  as  $N \rightarrow \infty$  and

$$\frac{1}{N} \sum_{i=1}^N \boldsymbol{\pi}_i - \boldsymbol{\pi} = \frac{1}{N} \sum_{i=1}^N \mathbf{v}_{\pi i} \xrightarrow{p} \mathbf{0}_{2k_x+1 \times 1}, \quad \text{as } N \rightarrow \infty. \quad (\text{A.74})$$

(A.73) and (A.74) establish (29), as desired.

Now suppose that the rank condition does not hold. Using model (1)-(2), vector of observations on the dependent variable,  $\mathbf{y}_i = (y_{i,p_T+1}, y_{i,p_T+2}, \dots, y_{i,T})'$ , can be written as (see (A.1))

$$\mathbf{y}_i = \mathbf{c}_{yi} + \Xi_i \boldsymbol{\pi}_i + \mathbf{F} \boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_i, \quad (\text{A.75})$$

where  $\mathbf{c}_{yi} = c_{yi} \boldsymbol{\tau}_{T-p_T}$  and  $\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_m)$  with  $\mathbf{f}_\ell = (f_{\ell,p_T+1}, f_{\ell,p_T+2}, \dots, f_{\ell,T})'$  for  $\ell = 1, 2, \dots, m$ . Substituting (A.75) into the definition of  $\hat{\boldsymbol{\pi}}_i$  in (26) and noting that  $\bar{\mathbf{M}}_q \mathbf{c}_{yi} = \mathbf{0}_{T-p_T \times 1}$  and  $(\Xi'_i \bar{\mathbf{M}}_q \Xi'_i)^{-1} \Xi'_i \bar{\mathbf{M}}_q \Xi_i \boldsymbol{\pi}_i = \boldsymbol{\pi}_i$ , we obtain the following expression for the mean group estimator,

$$\hat{\boldsymbol{\pi}}_{MG} = \frac{1}{N} \sum_{i=1}^N \boldsymbol{\pi}_i + \frac{1}{N} \sum_{i=1}^N \hat{\boldsymbol{\Psi}}_{\Xi, iT}^{-1} \frac{\Xi'_i \bar{\mathbf{M}}_q \boldsymbol{\varepsilon}_i}{T} + \frac{1}{N} \sum_{i=1}^N \hat{\boldsymbol{\Psi}}_{\Xi, iT}^{-1} \frac{\Xi'_i \bar{\mathbf{M}}_q \mathbf{F} \boldsymbol{\gamma}_i}{T}, \quad (\text{A.76})$$

where  $\hat{\boldsymbol{\Psi}}_{\Xi, iT}$  is defined in Assumption 7. Consider the asymptotics  $(N, T, p_T) \xrightarrow{j} \infty$ , such that  $p_T^3/T \rightarrow \varkappa$ , for some constant  $0 < \varkappa < \infty$ . The probability limit of the first term in (A.76) is established in (A.74). As before (see (A.71)),  $\hat{\boldsymbol{\Psi}}_{\Xi, iT}^{-1} = O_p(1)$  uniformly in  $i$  and using also (A.17) and (A.22) of Lemmas A.4 and A.6, respectively, we obtain

$$\frac{1}{N} \sum_{i=1}^N \hat{\boldsymbol{\Psi}}_{\Xi, iT}^{-1} \left( \frac{\Xi'_i \bar{\mathbf{M}}_q \boldsymbol{\varepsilon}_i}{T} \right) \xrightarrow{p} \mathbf{0}_{2k_x+1 \times 1}. \quad (\text{A.77})$$

Finally, consider the last term on the right side of (A.76). Since  $\boldsymbol{\Sigma}_{i\xi}$  is nonsingular, (A.12) of Lemma A.3 and (A.21) of Lemma A.6 establish that  $\hat{\boldsymbol{\Psi}}_{\Xi, iT}^{-1} \xrightarrow{p} \boldsymbol{\Sigma}_{i\xi}^{-1}$ , and together with (A.23) of Lemma A.6 we have

$$\frac{1}{N} \sum_{i=1}^N \hat{\boldsymbol{\Psi}}_{\Xi, iT}^{-1} \frac{\Xi'_i \bar{\mathbf{M}}_q \mathbf{F}}{T} \boldsymbol{\gamma}_i - \frac{1}{N} \sum_{i=1}^N \boldsymbol{\Sigma}_{i\xi}^{-1} \frac{\Xi'_i \mathbf{M}_h \mathbf{F}}{T} \boldsymbol{\gamma}_i \xrightarrow{p} \mathbf{0}_{2k_x+1 \times 1}.$$

Note that  $\boldsymbol{\gamma}_i = \boldsymbol{\eta}_{\gamma i} + (\bar{\boldsymbol{\gamma}}_w - \bar{\boldsymbol{\eta}}_{\gamma w}) \cdot \mathbf{F} (\bar{\boldsymbol{\gamma}}_w - \bar{\boldsymbol{\eta}}_{\gamma w})$  does not necessarily belong to the linear space spanned by the column vectors of  $\mathbf{Q}$  due to the truncation lag  $p_T$  and, in particular, we have  $T^{-1} \bar{\mathbf{M}}_h \mathbf{F} \bar{\boldsymbol{\gamma}}_w = O_p(\rho^{p_T})$ ,

$T^{-1}\bar{\mathbf{M}}_h\mathbf{F}\bar{\boldsymbol{\eta}}_{\gamma w} = O_p(N^{-1/2}\rho^{p_T})$ , and  $T^{-1}\boldsymbol{\Xi}'_i\bar{\mathbf{M}}_h\mathbf{F}\boldsymbol{\gamma}_i = T^{-1}\boldsymbol{\Xi}'_i\bar{\mathbf{M}}_h\mathbf{F}\boldsymbol{\eta}_{\gamma i} + O_p(N^{-1/2}\rho^{p_T}) + O_p(\rho^{p_T})$ , where  $\bar{\boldsymbol{\eta}}_{\gamma w} = O_p(N^{-1/2})$ ,  $|\rho| < 1$  and function  $\rho^\ell$ , for  $\ell = 1, 2, \dots$ , is an upper bound on the exponential decay of coefficients in the polynomial  $\boldsymbol{\Lambda}_w(L) = \sum_{i=1}^N w_i (\mathbf{I}_{k+1} - \mathbf{A}_i L)^{-1} \mathbf{A}_{0,i}^{-1} \mathbf{C}_i$  in the definition of  $\bar{\mathbf{Q}}_w$ . Now, when unobserved common factors are serially uncorrelated, we can use Lemma A.5 to obtain

$$\frac{1}{N} \sum_{i=1}^N \hat{\boldsymbol{\Psi}}_{\Xi, iT}^{-1} \left( \frac{\boldsymbol{\Xi}'_i \bar{\mathbf{M}}_q \mathbf{F}}{T} \right) \boldsymbol{\gamma}_i \xrightarrow{p} \mathbf{0}_{2k_x+1 \times 1}. \quad (\text{A.78})$$

Note that when factors are serially correlated and the rank condition does not hold then  $T^{-1}\boldsymbol{\Xi}'_i\bar{\mathbf{M}}_q\mathbf{F}\boldsymbol{\eta}_{\gamma i}$  does not converge to  $\mathbf{0}_{2k_x+1 \times 1}$  and as a result equation (A.78) would not hold. Using (A.74), (A.77) and (A.78) in (A.76) establish  $\hat{\boldsymbol{\pi}}_{MG} \rightarrow \boldsymbol{\pi}$ , when  $(N, T, p_T) \xrightarrow{j} \infty$  such that  $p_T^3/T \rightarrow \varkappa$  for some constant  $0 < \varkappa < \infty$ , as desired. ■

**Proof of Theorem 3.** Multiplying (A.76) by  $\sqrt{N}$  and substituting  $\boldsymbol{\pi}_i = \boldsymbol{\pi} + \mathbf{v}_{\pi i}$  we obtain

$$\sqrt{N}(\hat{\boldsymbol{\pi}}_{MG} - \boldsymbol{\pi}) = \frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{v}_{\pi i} + \frac{1}{\sqrt{N}} \sum_{i=1}^N \hat{\boldsymbol{\Psi}}_{\Xi, iT}^{-1} \frac{\boldsymbol{\Xi}'_i \bar{\mathbf{M}}_q \boldsymbol{\varepsilon}_i}{T} + \frac{1}{\sqrt{N}} \sum_{i=1}^N \hat{\boldsymbol{\Psi}}_{\Xi, iT}^{-1} \frac{\boldsymbol{\Xi}'_i \bar{\mathbf{M}}_q \mathbf{F} \boldsymbol{\gamma}_i}{T} \quad (\text{A.79})$$

where  $\hat{\boldsymbol{\Psi}}_{\Xi, iT}$  is defined in Assumption 7. Consider the asymptotics  $(N, T, p_T) \xrightarrow{j} \infty$  such that  $N/T \rightarrow \varkappa_1$  and  $p_T^3/T \rightarrow \varkappa_2$ , for some constants  $0 < \varkappa_1, \varkappa_2 < \infty$ . We establish convergence of the individual elements on the right side of (A.79) below.

It follows from (A.21) of Lemma A.6 and (A.12) of Lemma A.3 that

$$\hat{\boldsymbol{\Psi}}_{\Xi, iT} - \boldsymbol{\Sigma}_{i\xi} = o_p(N^{-1/2}) \text{ uniformly in } i. \quad (\text{A.80})$$

(A.80), (A.22) of Lemma A.6, and (A.26) of Lemma A.7 imply

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \hat{\boldsymbol{\Psi}}_{\Xi, iT}^{-1} \frac{\boldsymbol{\Xi}'_i \bar{\mathbf{M}}_q \boldsymbol{\varepsilon}_i}{T} \xrightarrow{p} \mathbf{0}_{2k_x+1 \times 1}. \quad (\text{A.81})$$

As in the proof of Theorem 2,  $\boldsymbol{\gamma}_i = \boldsymbol{\eta}_{\gamma i} + (\bar{\boldsymbol{\gamma}}_w - \bar{\boldsymbol{\eta}}_{\gamma w})$ ,  $\mathbf{F}(\bar{\boldsymbol{\gamma}}_w - \bar{\boldsymbol{\eta}}_{\gamma w})$  does not necessarily belong to the linear space spanned by the column vectors of  $\mathbf{Q}$  due to the truncation lag  $p_T$  and, in particular, we have  $T^{-1}\boldsymbol{\Xi}'_i\bar{\mathbf{M}}_h\mathbf{F}\boldsymbol{\gamma}_i = T^{-1}\boldsymbol{\Xi}'_i\bar{\mathbf{M}}_h\mathbf{F}\boldsymbol{\eta}_{\gamma i} + O_p(N^{-1/2}\rho^{p_T}) + O_p(\rho^{p_T})$ , where  $\bar{\boldsymbol{\eta}}_{\gamma w} = O_p(N^{-1/2})$ ,  $|\rho| < 1$  and function  $\rho^\ell$ , for  $\ell = 1, 2, \dots$ , is an upper bound on the exponential decay of coefficients in the polynomial  $\boldsymbol{\Lambda}_w(L) = \sum_{i=1}^N w_i (\mathbf{I}_{k+1} - \mathbf{A}_i L)^{-1} \mathbf{A}_{0,i}^{-1} \mathbf{C}_i$  in the definition of  $\bar{\mathbf{Q}}_w$ . Using now (A.21) and (A.23) of Lemma A.6 and noting that  $\sqrt{N}\rho^{p_T} \rightarrow 0$  yields

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \hat{\boldsymbol{\Psi}}_{\Xi, iT}^{-1} \frac{\boldsymbol{\Xi}'_i \bar{\mathbf{M}}_q \mathbf{F}}{T} \boldsymbol{\gamma}_i - \frac{1}{\sqrt{N}} \sum_{i=1}^N \left( \frac{\boldsymbol{\Xi}'_i \mathbf{M}_h \boldsymbol{\Xi}_i}{T} \right)^{-1} \frac{\boldsymbol{\Xi}'_i \mathbf{M}_h \mathbf{F}}{T} \boldsymbol{\eta}_{\gamma i} \xrightarrow{p} \mathbf{0}_{2k_x+1 \times 1}. \quad (\text{A.82})$$

Using (A.81)-(A.82) in (A.79), we obtain

$$\sqrt{N}(\hat{\boldsymbol{\pi}}_{MG} - \boldsymbol{\pi}) \stackrel{d}{\sim} \boldsymbol{\vartheta}_{\pi i},$$

where

$$\boldsymbol{\vartheta}_{\pi i} = \frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{v}_i + \frac{1}{\sqrt{N}} \sum_{i=1}^N \left( \frac{\boldsymbol{\Xi}'_i \mathbf{M}_h \boldsymbol{\Xi}_i}{T} \right)^{-1} \frac{\boldsymbol{\Xi}'_i \mathbf{M}_h \mathbf{F}}{T} \boldsymbol{\eta}_{\gamma i}, \quad (\text{A.83})$$

and recall that  $\mathbf{v}_i$  and  $\boldsymbol{\eta}_{\gamma i}$  are independently distributed across  $i$ . It now follows that  $\sqrt{N}(\hat{\boldsymbol{\pi}}_{MG} - \boldsymbol{\pi}) \rightarrow$

$N \begin{pmatrix} \mathbf{0} \\ 2k_x+1 \times 1 \end{pmatrix}, \Sigma_{MG}$ , where

$$\Sigma_{MG} = \Omega_\pi + \lim_{N \rightarrow \infty} \left[ \frac{1}{N} \sum_{i=1}^N \Sigma_{i\xi}^{-1} \mathbf{Q}_{if} \Omega_\gamma \mathbf{Q}_{if}' \Sigma_{i\xi}^{-1} \right], \quad (\text{A.84})$$

in which  $\Omega_\pi = \text{Var}(\pi_i) = \text{Var}(\mathbf{v}_{\pi i})$ ,  $\Omega_\gamma = \text{Var}(\gamma_i) = \text{Var}(\eta_{\gamma i})$ , and  $\Sigma_{i\xi} = p \lim T^{-1} \Xi_i' \mathbf{M}_h \Xi_i$  and  $\mathbf{Q}_{if} = p \lim T^{-1} \Xi_i' \mathbf{M}_h \mathbf{F}$  are defined by (A.12) and (A.13) of Lemma A.3, respectively. When the rank condition stated in Assumption 6 hold then  $\mathbf{Q}_{if} = \begin{pmatrix} \mathbf{0} \\ 2k_x+1 \times m \end{pmatrix}$  and  $\Sigma_{MG}$  reduces to  $\Sigma_{MG} = \Omega_\pi$ .

Consider now the non-parametric variance estimator (32) and the same assumptions on the divergence of  $(N, T, p_T)$ . We have

$$\hat{\pi}_i - \hat{\pi}_{MG} = (\hat{\pi}_i - \pi) + (\pi - \hat{\pi}_{MG}),$$

where  $\sqrt{N}(\pi - \hat{\pi}_{MG}) \xrightarrow{d} N \begin{pmatrix} \mathbf{0} \\ 2k_x+1 \times 1 \end{pmatrix}, \Sigma_{MG}$  with  $\|\Sigma_{MG}\| < K$ . It therefore follows that

$$\frac{1}{N-1} \sum_{i=1}^N (\hat{\pi}_i - \hat{\pi}_{MG}) (\hat{\pi}_i - \hat{\pi}_{MG})' = \frac{1}{N-1} \sum_{i=1}^N (\hat{\pi}_i - \pi) (\hat{\pi}_i - \pi)' + O_p(N^{-1/2}).$$

Consider now  $\hat{\pi}_i - \pi$ . As before, using the definition of  $\pi_i$  in (26) and substituting  $\pi_i = \pi + \mathbf{v}_{\pi i}$  we obtain

$$\hat{\pi}_i - \pi = \mathbf{v}_{\pi i} + \hat{\Psi}_{\Xi, iT}^{-1} \frac{\Xi_i' \bar{\mathbf{M}}_q \varepsilon_i}{T} + \hat{\Psi}_{\Xi, iT}^{-1} \frac{\Xi_i' \bar{\mathbf{M}}_q \mathbf{F} \gamma_i}{T}.$$

Using (A.81)-(A.82), we have

$$\begin{aligned} \frac{1}{N-1} \sum_{i=1}^N (\hat{\pi}_i - \pi)' (\hat{\pi}_i - \pi) &= \frac{1}{N-1} \sum_{i=1}^N \mathbf{v}_{\pi i} \mathbf{v}_{\pi i}' \\ &+ \frac{1}{N-1} \sum_{i=1}^N \left( \frac{\Xi_i' \mathbf{M}_h \Xi_i}{T} \right)^{-1} \frac{\Xi_i' \mathbf{M}_h \mathbf{F}}{T} \eta_{\gamma i} \eta_{\gamma i}' \left( \frac{\Xi_i' \mathbf{M}_h \mathbf{F}}{T} \right)' \left( \frac{\Xi_i' \mathbf{M}_h \Xi_i}{T} \right)^{-1} \\ &+ o_p(1) \\ &= \frac{1}{N-1} \sum_{i=1}^N \mathbf{v}_{\pi i} \mathbf{v}_{\pi i}' + \frac{1}{N-1} \sum_{i=1}^N \Sigma_{i\xi}^{-1} \mathbf{Q}_{if} \eta_{\gamma i} \eta_{\gamma i}' \mathbf{Q}_{if}' \Sigma_{i\xi}^{-1} + o_p(1), \end{aligned}$$

where  $\Sigma_{i\xi} = p \lim T^{-1} \Xi_i' \mathbf{M}_h \Xi_i$  and  $\mathbf{Q}_{if} = p \lim T^{-1} \Xi_i' \mathbf{M}_h \mathbf{F}$  are defined by (A.12) and (A.13) of Lemma A.3, respectively. Note that  $\mathbf{v}_{\pi i}$  and  $\eta_{\gamma i}$  are independently distributed across  $i$  and therefore  $\frac{1}{N-1} \sum_{i=1}^N (\hat{\pi}_i - \pi)' (\hat{\pi}_i - \pi) - \Sigma_{MG} \xrightarrow{p} 0$  and  $\hat{\Sigma}_{MG} \xrightarrow{p} \Sigma_{MG}$ , as required. ■

## References

- Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica* 77, 1229–1279.
- Bai, J. and S. Ng (2007). Determining the number of primitive shocks in factor models. *Journal of Business and Economic Statistics* 25, 52–60.
- Berk, K. N. (1974). Consistent autoregressive spectral estimates. *The Annals of Statistics* 2, 489–502.
- Bruno, G. S. (2005). Approximating the bias of the LSDV estimator for dynamic unbalanced panel data models. *Economics Letters* 87, 361–366.
- Bun, M. J. G. (2003). Bias correction in the dynamic panel data model with a nonscalar disturbance covariance matrix. *Econometric Reviews* 22, 29–58.
- Bun, M. J. G. and M. A. Carree (2005). Bias-corrected estimation in dynamic panel data models. *Journal of Business and Economic Statistics* 23, 200–210.
- Bun, M. J. G. and M. A. Carree (2006). Bias-corrected estimation in dynamic panel data models with heteroscedasticity. *Economics Letters* 92, 220–227.
- Bun, M. J. G. and J. Kiviet (2003). On the diminishing returns of higher order terms in asymptotic expansions of bias. *Economic Letters* 19, 145–152.
- Canova, F. and M. Ciccarelli (2004). Forecasting and turning point predictions in a Bayesian panel VAR model. *Journal of Econometrics* 120, 327–359.
- Canova, F. and M. Ciccarelli (2009). Estimating multicountry VAR models. *International Economic Review* 50, 929–959.
- Canova, F. and A. Marcet (1999). The poor stay poor: Non-convergence across countries and regions. Mimeo, June 1999.
- Chudik, A. and M. H. Pesaran (2011). Infinite dimensional VARs and factor models. *Journal of Econometrics* 163, 4–22.
- Chudik, A. and M. H. Pesaran (2013a). Aggregation in large dynamic panels. *forthcoming in Journal of Econometrics*.
- Chudik, A. and M. H. Pesaran (2013b). Econometric analysis of high dimensional VARs featuring a dominant unit. *Econometric Reviews* 32, 592–649.
- Chudik, A., M. H. Pesaran, and E. Tosetti (2011). Weak and strong cross section dependence and estimation of large panels. *Econometrics Journal* 14, C45–C90.
- Davidson, J. (1994). *Stochastic Limit Theory*. Oxford University Press.
- Dhaene, G. and K. Jochmans (2012). Split-panel jackknife estimation of fixed-effect models. Mimeo, 21 July 2012.
- Everaert, G. and T. D. Groote (2012). Common correlated effects estimation of dynamic panels with cross-sectional dependence. Mimeo, 9 November 2012.
- Everaert, G. and L. Ponzi (2007). Bootstrap-based bias correction for dynamic panels. *Journal of Economic Dynamics and Control* 31, 1160–1184.
- Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2005). The generalized dynamic factor model: One-sided estimation and forecasting. *Journal of the American Statistical Association* 100, 830–840.
- Garcia-Ferrer, A., R. A. Highfield, F. Palm, and A. Zellner (1987). Macroeconomic forecasting using pooled international data. *Journal of Business and Economic Statistics* 5, 53–67.

- Giannone, D., L. Reichlin, and L. Sala (2005). Monetary policy in real time. In M. Gertler and K. Rogoff (Eds.), *NBER Macroeconomics Annual 2004*, Volume 19, pp. 161–200. MIT Press.
- Hahn, J. and G. Kuersteiner (2002). Asymptotically unbiased inference for a dynamic panel model with fixed effects when both N and T are large. *Econometrica* 70, 1639–1657.
- Hahn, J. and G. Kuersteiner (2011). Bias reduction for dynamic nonlinear panel models with fixed effects. *Econometric Theory* 27, 1152–1191.
- Hahn, J. and H. Moon (2006). Reducing bias of MLE in a dynamic panel model. *Econometric Theory* 22, 499–512.
- Hahn, J. and W. Newey (2004). Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica* 72, 1295–1319.
- Hsiao, C., M. H. Pesaran, and A. K. Tahmiscioglu (1999). Bayes estimation of short-run coefficients in dynamic panel data models. In C. Hsiao, K. Lahiri, L.-F. Lee, and M. H. Pesaran (Eds.), *Analysis of Panels and Limited Dependent Variables: A Volume in Honour of G. S. Maddala*, Chapter 11, pp. 268–296. Cambridge University Press.
- Hurwicz, L. (1950). Least squares bias in time series. In T. C. Koopman (Ed.), *Statistical Inference in Dynamic Economic Models*, pp. 365–383. New York: Wiley.
- Kapetanios, G., M. H. Pesaran, and T. Yagamata (2011). Panels with nonstationary multifactor error structures. *Journal of Econometrics* 160, 326–348.
- Kiviet, J. F. (1995). On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. *Journal of Econometrics* 68, 53–78.
- Kiviet, J. F. (1999). Expectations of expansions for estimators in a dynamic panel data model; some results for weakly-exogenous regressors. In C. Hsiao, K. Lahiri, L.-F. Lee, and M. H. Pesaran (Eds.), *Analysis of Panel Data and Limited Dependent Variables*. Cambridge University Press, Cambridge.
- Kiviet, J. F. and G. D. A. Phillips (1993). Alternative bias approximations in regressions with a lagged-dependent variable. *Econometric Theory* 9, 62–80.
- Lee, N., H. R. Moon, and M. Weidner (2011). Analysis of interactive fixed effects dynamic linear panel regression with measurement error. Cemmap working paper CWP37/11.
- Mark, N. C. and D. Sul (2003). Cointegration vector estimation by panel DOLS and long-run money demand. *Oxford Bulletin of Economics and Statistics* 65, 655–680.
- Moon, H. R. and M. Weidner (2010a). Dynamic linear panel regression models with interactive fixed effects. Mimeo, July 2010.
- Moon, H. R. and M. Weidner (2010b). Linear regression for panel with unknown number of factors as interactive fixed effects. Mimeo, July 2010.
- Newey, W. K. and R. J. Smith (2004). Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica* 72, 219–255.
- Pedroni, P. (2000). Fully modified OLS for heterogeneous cointegrated panels. *Advances in Econometrics* 15, 93–130.
- Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with multifactor error structure. *Econometrica* 74, 967–1012.
- Pesaran, M. H., Y. Shin, and R. P. Smith (1999). Pooled mean group estimation of dynamic heterogeneous panels. *Journal of the American Statistical Association* 94, 621–634.
- Pesaran, M. H., L. V. Smith, and T. Yamagata (2013). A panel unit root test in the presence of a multifactor error structure. *forthcoming in Journal of Econometrics*.

- Pesaran, M. H. and R. Smith (1995). Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics* 68, 79–113.
- Pesaran, M. H. and E. Tosetti (2011). Large panels with common factors and spatial correlations. *Journal of Econometrics* 161, 182–202.
- Pesaran, M. H. and Z. Zhao (1999). Bias reduction in estimating long-run relationships from dynamic heterogeneous panels. In C. Hsiao, K. Lahiri, L.-F. Lee, and M. H. Pesaran (Eds.), *Analysis of Panels and Limited Dependent Variables: A Volume in Honour of G. S. Maddala*, Chapter 12, pp. 297–322. Cambridge University Press.
- Phillips, P. C. B. and D. Sul (2003). Dynamic panel estimation and homogeneity testing under cross section dependence. *Econometrics Journal* 6, 217–259.
- Phillips, P. C. B. and D. Sul (2007). Bias in dynamic panel estimation with fixed effects, incidental trends and cross section dependence. *Journal of Econometrics* 137, 162–188.
- Rudin, W. (1987). *Real and Complex Analysis*. McGraw-Hill.
- Said, E. and D. A. Dickey (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* 71, 599–607.
- So, B. S. and D. W. Shin (1999). Recursive mean adjustment in time series inferences. *Statistics & Probability Letters* 43, 65–73.
- Song, M. (2013). Asymptotic theory for dynamic heterogeneous panels with cross-sectional dependence and its applications. Mimeo, January 2013.
- Stock, J. H. and M. W. Watson (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics* 20, 147–162.
- Stock, J. H. and M. W. Watson (2005). Implications of dynamic factor models for VAR analysis. NBER Working Paper No. 11467.
- Zellner, A. and C. Hong (1989). Forecasting international growth rates using Bayesian shrinkage and other procedures. *Journal of Econometrics* 40, 183–202.
- Zellner, A., C. Hong, and C. ki Min (1991). Forecasting turning points in international output growth rates using Bayesian exponentially weighted autoregression, time-varying parameter, and pooling techniques. *Journal of Econometrics* 49, 275–304.
- Zhang, P. and D. Small (2006). Bayesian inference for random coefficient dynamic panel data models. Mimeo, 20 February 2006.